

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2004, 07:36:16 ; Search time 20.9028 Seconds  
(without alignments)  
593.639 Million cell updates/sec

Title: US-09-805-290A-18

Perfect score: 680  
Sequence: 1 QVQLQDSGGGLVQAGGSLRL.....WGQGTQVTVSSAHSEDPSS 129

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : PIR.78:\*  
1: PIR1:\*  
2: PIR2:\*  
3: PIR3:\*  
4: PIR4:\*

Fred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	413.5	60.8	119	2	IG heavy chain V r
2	409.5	60.2	119	2	IG heavy chain V r
3	409	60.1	120	2	IG heavy chain V r
4	408.5	60.1	121	2	IG heavy chain V r
5	408.5	60.1	123	2	IG heavy chain - h
6	407.5	59.9	125	2	IG heavy chain V r
7	403.5	59.3	121	2	IG heavy chain V r
8	402.5	59.2	119	2	IG heavy chain - h
9	402	59.1	114	2	IG heavy chain V r
10	401.5	59.0	138	2	IG heavy chain V r
11	401	59.0	143	2	IG heavy chain V r
12	400.5	58.9	140	2	IG heavy chain V r
13	400	58.8	116	2	IG heavy chain - h
14	398.5	58.6	119	2	IG heavy chain - h
15	398	58.5	114	2	IG heavy chain V r
16	396.5	58.3	117	2	IG heavy chain V r
17	396.5	58.3	121	2	IG heavy chain - h
18	395.5	58.2	121	2	IG heavy chain - h
19	395.5	58.2	123	2	IG heavy chain V r
20	394.5	58.0	127	2	IG heavy chain - h
21	394.5	58.0	134	2	IG heavy chain V r
22	394.5	58.0	139	2	IG heavy chain V r
23	393.5	57.9	121	2	IG heavy chain V r
24	393	57.8	147	2	IG heavy chain V r
25	392.5	57.7	120	2	IG heavy chain V r
26	391.5	57.6	140	2	IG heavy chain V r
27	390	57.4	114	2	IG heavy chain - h
28	389.5	57.3	120	2	IG heavy chain V r
29	389.5	57.3	123	2	IG heavy chain V r

30	389.5	57.3	135	2	IG heavy chain V r
31	389	57.2	122	2	IG heavy chain V r
32	388	57.1	140	2	IG heavy chain pre
33	387.5	57.0	121	2	IG heavy chain - h
34	387.5	57.0	123	2	IG heavy chain - h
35	387.5	57.0	134	2	IG heavy chain V r
36	387	56.9	124	2	IG heavy chain V r
37	387	56.9	254	2	IG heavy chain V r
38	386.5	56.8	117	2	IG heavy chain V r
39	386	56.8	122	2	IG heavy chain V r
40	385.5	56.7	120	2	IG heavy chain V r
41	385.5	56.7	121	2	IG heavy chain V r
42	385.5	56.7	160	2	IG heavy chain - h
43	385	56.6	151	2	IG heavy chain pre
44	384.5	56.5	117	2	IG heavy chain V r
45	384	56.5	114	2	IG heavy chain V r

## ALIGNMENTS

## RESULT 1

C36005

IG heavy chain V region (30p1) - human

C/Species: Homo sapiens (man)

C/Date: 21-Dec-1990 #sequence\_revision 21-Dec-1990 #text\_change 16-Aug-1996

C/Accession: C36005

R/Schroeder Jr., H.W.; Wang, J.Y.

Proc. Natl. Acad. Sci. U.S.A. 87, 6146-6150, 1990

A/Title: Preferential utilization of conserved immunoglobulin heavy chain variable gene ;

A/Reference number: A36005; M01D:90349571; PMID:2117273

A/Status: preliminary

A/Molecule type: mRNA

A/Residues: 1-119 <SCH>

A/Cross-references: GB:M18513

C/Superfamily: immunoglobulin V region; immunoglobulin homology

C/Keywords: heterotrimer; immunoglobulin

F/15-98/Domain: immunoglobulin homology <IMM>

Query Match Best Local Similarity 60.8%; Score 413.5; DB 2; Length 119;  
Matches 87; Conservative 8; Mismatches 20; Indels 9; Gaps 3;

QY	1	QVQLQDSGGGLVQAGGSLRLSCAASGIFSSDLMGWIRQAPKEREAVARIT-RCGTTSY	59
DB	1	EVQLLESGGGLVQPGSRLSCASGFTFSYAMSWVRQAPGKGLIEWSAISGSGSTYY	60
QY	60	ADSVKGRFTISRDNAMNTMYLQNNSLKPEDTAVYYCCNARRSNVDRSWG---DYWGQGTQV	116
DB	61	ADSVKGRFTISRDNKNTLYLQNNSLRAEDTVYCAK-----DAGKSGGPDYWGQGLV	115
QY	117	TVSS 120	
DB	116	TVSS 119	

## RESULT 2

D36005

IG heavy chain V region (M43) - human

C/Species: Homo sapiens (man)

C/Date: 21-Dec-1990 #sequence\_revision 21-Dec-1990 #text\_change 16-Dec-1996

C/Accession: D36005

R/Schroeder Jr., H.W.; Wang, J.Y.

A/Title: Preferential utilization of conserved immunoglobulin heavy chain variable gene ;

A/Reference number: A36005; M01D:90349571; PMID:2117273

A/Status: preliminary

A/Molecule type: mRNA

A/Residues: 1-119 <SCH>

A/Cross-references: GB:M34024

C/Genetics:





A:Accession: S30531  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-125 <MAR>  
 A:Cross-references: EMBL:Z18317  
 C:Superfamily: immunoglobulin V region; immunoglobulin homology  
 C:Keywords: heterotrimer; immunoglobulin  
 F:15-98/Domain: immunoglobulin homology <IMM>

Query Match 59.9%; Score 407.5; DB 2; Length 125;  
 Best Local Similarity 69.0%; Pred. No. 1.7e-29;  
 Matches 87; Conservative 10; Mismatches 22; Indels 7; Gaps 3;

QY 1 QVQLDSSGGGLVQAGSLRLSCAASGSIFFSDLMGWYRQAPGKERAVARITRGCTT-SY 59  
 DB 1 EVQLVESGGGLVQPGGSLRLSCAASGFTFSYSNMWRQAPGKLEWISYISSSSSTIYY 60  
 QY 60 ADSVKGRTISRDNKNTMYLQMSLKPEDTAVYYCNARRSNYDRS-----WGPWGQGT 114  
 DB 61 ADSVKGRTISRDNKNTMYLQMSLKRAEDTAVYYC-ARRSNYSSSGYSHYFDYWGQGT 119  
 QY 115 QVTWS 120  
 DB 120 LVTWS 125

## RESULT 7

S26798  
 Ig heavy chain V region - human  
 C:Species: Homo sapiens (man)  
 C>Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 20-Jun-2000  
 C:Accession: S26798  
 R:Morlati, F.; Newton, J.A.; Wang, J.Y.; Schroeder Jr., H.W.  
 Eur. J. Immunol. 22, 241-245, 1992  
 A:Title: The human cord blood antibody repertoire. Frequent usage of the V(H)7 gene faml  
 A:Reference number: S26786; MUID:92111632; PMID:1730251  
 A:Accession: S26798  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-121 <MOR>  
 A:Cross-references: EMBL:X61015; NID:932795; PIND:CAA3349.1; PID:G1335126  
 C:Superfamily: immunoglobulin V region; immunoglobulin homology  
 C:Keywords: heterotrimer; immunoglobulin  
 F:15-98/Domain: immunoglobulin homology <IMM>

Query Match 59.3%; Score 403.5; DB 2; Length 121;  
 Best Local Similarity 69.4%; Pred. No. 3.8e-29;  
 Matches 84; Conservative 8; Mismatches 28; Indels 1; Gaps 1;

QY 1 QVQLDSSGGGLVQAGSLRLSCAASGSIFFSDLMGWYRQAPGKERAVARITRGCTT-SY 59  
 DB 1 EVQLVESGGGLVQPGGSLRLSCAASGFTFSYSNMWRQAPGKLEWISYISSSSSTIYY 60  
 QY 60 ADSVKGRTISRDNKNTMYLQMSLKPEDTAVYYCNARRSNYDRSMDYWGQGTQVTVS 119  
 DB 61 ADSVKGRTISRDNKNTMYLQMSLKPEDTAVYYC-ARRSNYSSSGYSHYFDYWGQGLTVTS 120  
 QY 120 S 120  
 DB 121 S 121

## RESULT 8

S31108  
 Ig heavy chain - human  
 C:Species: Homo sapiens (man)  
 C>Date: 02-Dec-1993 #sequence\_revision 26-May-1995 #text\_change 17-Mar-1999  
 C:Accession: S31108  
 R:Raapport, F.M.; Timmers, E.; Kenter, M.J.H.; van Tol, M.J.D.; Vossen, J.M.; Schuurman  
 Eur. J. Immunol. 22, 247-251, 1992  
 A:Title: Restricted utilization of germ-line V(H)3 genes and short diverse third complen  
 A:Reference number: S31104; MUID:92111633; PMID:1730252  
 A:Accession: S31108

A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: mRNA  
 A:Residues: 1-119 <RAA>  
 A:Cross-references: EMBL:X62956  
 A>Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1991  
 C:Superfamily: immunoglobulin V region; immunoglobulin homology  
 C:Keywords: heterotrimer; immunoglobulin  
 F:15-98/Domain: immunoglobulin homology <IMM>

Query Match 59.2%; Score 402.5; DB 2; Length 119;  
 Best Local Similarity 70.7%; Pred. No. 4.6e-29;  
 Matches 87; Conservative 8; Mismatches 21; Indels 7; Gaps 3;

QY 1 QVQLDSSGGGLVQAGSLRLSCAASGSIFFSDLMGWYRQAPGKERAVARIT-RGGTTSY 59  
 DB 1 EVQLVESGGGLVQPGGSLRLSCAASGFTFSYAMSVWRQAPGKLEWISYISSSSSTIYY 60  
 QY 60 ADSVKGRTISRDNKNTMYLQMSLKPEDTAVYYCNARRSNYDRSG--DIWGQGTQVT 117  
 DB 61 ADSVKGRTISRDNKNTMYLQMSLKRAEDTAVYYCANDR---RLTGFDYWGQGLTVT 116  
 QY 118 VSS 120  
 DB 117 VSS 119

## RESULT 9

S46390  
 Ig heavy chain V region - human  
 C:Species: Homo sapiens (man)  
 C>Date: 27-Jan-1995 #sequence\_revision 27-Jan-1995 #text\_change 20-Jun-2000  
 C:Accession: S46390  
 R:Frigit, M.; Marks, J.D.; Winter, G.; Griffiths, A.D.  
 U. Mol. Biol. 239, 68-78, 1994  
 A:Title: In vitro assembly of repertoires of antibody chains on the surface of phage by  
 A:Reference number: S46390; MUID:94254092; PMID:196048  
 A:Accession: S46390  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-114 <FTG>  
 A:Cross-references: EMBL:Z31686; NID:G509782; PIND:CAA83491.1; PID:G1335143  
 C:Superfamily: immunoglobulin V region; immunoglobulin homology  
 C:Keywords: heterotrimer; immunoglobulin  
 F:15-98/Domain: immunoglobulin homology <IMM>

Query Match 59.1%; Score 402; DB 2; Length 114;  
 Best Local Similarity 69.4%; Pred. No. 4.9e-29;  
 Matches 84; Conservative 8; Mismatches 21; Indels 8; Gaps 2;

QY 1 QVQLDSSGGGLVQAGSLRLSCAASGSIFFSDLMGWYRQAPGKERAVARITRGCTT-SY 59  
 DB 1 EVQLVESGGGLVQPGGSLRLSCAASGFTFSYAMSVWRQAPGKLEWISYISSSSSTIYY 60  
 QY 60 ADSVKGRTISRDNKNTMYLQMSLKPEDTAVYYCNARRSNYDRSMDYWGQGTQVTVS 119  
 DB 61 ADSVKGRTISRDNKNTMYLQMSLKRAEDTAVYYC-----ARDYDYGQGLTVTS 113  
 QY 120 S 120  
 DB 114 S 114

## RESULT 10

S31666  
 Ig heavy chain V region - human (fragment)  
 C:Species: Homo sapiens (man)  
 C>Date: 22-Nov-1993 #sequence\_revision 10-Nov-1995 #text\_change 23-Jul-1999  
 C:Accession: S31666  
 R:Cuisinier, A.M.; Gauthier, L.; Boubli, L.; Fougereau, M.; Tonnelle, C.  
 submitted to the EMBL Data Library, June 1992  
 A:Description: Mechanisms that generate human immunoglobulin diversity operate from the  
 A:Reference number: S31585  
 A:Accession: S31666

A/Status: preliminary  
A/Molecule type: mRNA  
A/Residues: 1-138 <CUI>  
A/Cross-references: EMBL:Z14202; NID:g30963; PIDD:CAA78571.1; PID:g30964  
C/Superfamily: immunoglobulin V region; immunoglobulin homology  
C/Keywords: heterotetramer; immunoglobulin  
F/34-117/Domain: immunoglobulin homology <IMM>

Query Match 59.0%; Score 401.5; DB 2; Length 138;  
Best Local Similarity 69.7%; Pred. No. 6.6e-29;  
Matches 85; Conservative 10; Mismatches 22; Indels 5; Gaps 3;

QY 1 QVQLDSGGGLVQAGSLRLSCAASGIFSSDLMGMYRQAPGKEREAARITRGTT-SY 59  
DB 20 EVQLVESGGGLVQPGGSLRLSCAASGFTSSYAMSWVRQAPGKLEWVSATISGSGSTYY 79  
QY 60 ADSVKGRFTISRDNKNTMYLQNSLKPEDTAVYYCNARRSNYDRSM-CDYWGQGTQVTV 118  
DB 80 ADSVKGRFTISRDNKNTMYLQNSLKPEDTAVYYCNARRSNYDRSM-CDYWGQGTQVTV 136  
QY 119 SS 120  
DB 137 SS 138

## RESULT 11

S23624  
Ig heavy chain V region - human (fragment)

C/Species: Homo sapiens (man)  
C/Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 23-Jul-1999  
C/Accession: S23624

R/Olee, T.; Lu, E.W.; Huang, D.F.; Soto-Gil, R.W.; Deflos, M.; Kozin, F.; Carson, D.A.;  
J. Exp. Med. 175, 831-842, 1992  
A/Title: Genetic analysis of self-associating immunoglobulin G rheumatoid factors from

A/Reference number: S23623; MUID:92156804; PMID:1740665

A/Accession: S23624

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-143 <OLE>

A/Cross-references: EMBL:X62973; NID:g32012; PIDD:CAA42224.1; PID:g32013

C/Superfamily: immunoglobulin V region; immunoglobulin homology

C/Keywords: heterotetramer; immunoglobulin  
F/15-98/Domain: immunoglobulin homology <IMM>

Query Match 59.0%; Score 401; DB 2; Length 143;  
Best Local Similarity 68.2%; Pred. No. 7.6e-29;  
Matches 88; Conservative 11; Mismatches 24; Indels 6; Gaps 4;

QY 1 QVQLDSGGGLVQAGSLRLSCAASGIFSSDLMGMYRQAPGKEREAARITRGTT-SY 59  
DB 1 EVQLVESGGGLVQPGGSLRLSCAASGFTSSYAMSWVRQAPGKLEWVSATISGSGSTYY 60  
QY 60 ADSVKGRFTISRDNKNTMYLQNSLKPEDTAVYYCNARRSNYDRSMGDIYGGGTQVTV 119  
DB 61 ADSVKGRFTISRDNKNTMYLQNSLKPEDTAVYYCNARRSNYDRSMGDIYGGGTQVTV 116  
QY 120 SAHSHSDPS 128  
DB 117 SA-STKGPS 124

## RESULT 12

S31588  
Ig heavy chain V region - human (fragment)

C/Species: Homo sapiens (man)  
C/Date: 22-Nov-1993 #sequence\_revision 10-Nov-1995 #text\_change 23-Jul-1999  
C/Accession: S31588

R/Ciesnier, A.M.; Gauchier, L.; Boudit, L.; Fougereau, M.; Tonnelle, C.

submitted to the EMBL Data Library, June 1992

A/Description: Mechanisms that generate human immunoglobulin diversity operate from the

A/Reference number: S31585

A/Accession: S31588

A/Status: preliminary

A/Molecule type: mRNA  
A/Residues: 1-140 <CUI>  
A/Cross-references: EMBL:Z14200; NID:g30957; PIDD:CAA78569.1; PID:g30958  
C/Superfamily: immunoglobulin V region; immunoglobulin homology  
C/Keywords: heterotetramer; immunoglobulin  
F/34-117/Domain: immunoglobulin homology <IMM>

Query Match 58.9%; Score 400.5; DB 2; Length 140;  
Best Local Similarity 68.6%; Pred. No. 8.3e-29;  
Matches 63; Conservative 11; Mismatches 26; Indels 1; Gaps 1;

QY 1 QVQLDSGGGLVQAGSLRLSCAASGIFSSDLMGMYRQAPGKEREAARITRGTT-SY 59  
DB 20 EVQLVESGGGLVQPGGSLRLSCAASGFTSSYAMSWVRQAPGKLEWVSATISGSGSTYY 79  
QY 60 ADSVKGRFTISRDNKNTMYLQNSLKPEDTAVYYCNARRSNYDRSMGDIYGGGTQVTV 119  
DB 80 ADSVKGRFTISRDNKNTMYLQNSLKPEDTAVYYCNARRSNYDRSMGDIYGGGTQVTV 139  
QY 120 S 120  
DB 140 S 140

## RESULT 13

S3110  
Ig heavy chain - human

C/Species: Homo sapiens (man)  
C/Date: 02-Dec-1993 #sequence\_revision 26-May-1995 #text\_change 17-Mar-1999  
C/Accession: S3110

R/Rapaport, F.M.; Timmers, E.; Kenter, M.J.H.; van Tol, M.J.D.; Vossen, J.M.; Schuurman,

Eur. J. Immunol. 22, 247-251, 1992

A/Title: Restricted utilization of germ-line V(H)3 genes and short diverse third compleme

A/Reference number: S31104; MUID:92111633; PMID:1730252

A/Accession: S3110

A/Status: preliminary; nucleic acid sequence not shown; translation not shown

A/Molecule type: mRNA

A/Residues: 1-116 <RAA>

A/Cross-references: EMBL:X62958

A/Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1991

C/Superfamily: immunoglobulin V region; immunoglobulin homology

C/Keywords: heterotetramer; immunoglobulin  
F/15-97/Domain: immunoglobulin homology <IMM>

Query Match 58.8%; Score 400; DB 2; Length 116;  
Best Local Similarity 69.2%; Pred. No. 7.5e-29;  
Matches 83; Conservative 11; Mismatches 22; Indels 4; Gaps 2;

QY 1 QVQLDSGGGLVQAGSLRLSCAASGIFSSDLMGMYRQAPGKEREAARITRGTT-SY 60  
DB 1 EVQLVETGGGLVQPGGSLRLSCAASGFTSSYAMSWVRQAPGKLEWVSATISGSGSTYY 60  
QY 61 DSVKGRFTISRDNKNTMYLQNSLKPEDTAVYYCNARRSNYDRSMGDIYGGGTQVTV 120  
DB 61 DSVKGRFTISRDNKNTMYLQNSLKPEDTAVYYCNARRSNYDRSMGDIYGGGTQVTV 116

## RESULT 14

S31107  
Ig heavy chain - human

C/Species: Homo sapiens (man)  
C/Date: 02-Dec-1993 #sequence\_revision 26-May-1995 #text\_change 17-Mar-1999  
C/Accession: S31107

R/Rapaport, F.M.; Timmers, E.; Kenter, M.J.H.; van Tol, M.J.D.; Vossen, J.M.; Schuurman,

Eur. J. Immunol. 22, 247-251, 1992

A/Title: Restricted utilization of germ-line V(H)3 genes and short diverse third compleme

A/Reference number: S31104; MUID:92111633; PMID:1730252

A/Accession: S31107

A/Status: preliminary; nucleic acid sequence not shown; translation not shown

A/Molecule type: mRNA

A/Residues: 1-119 <RAA>

A/Cross-references: EMBL:X62955

A/Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1991

C:Superfamily: immunoglobulin V region; immunoglobulin homology  
C:Keywords: heterotetramer; immunoglobulin  
F:15-98/Domain: immunoglobulin homology <IMM>

Query Match 58.6%; Score 398.5; DB 2; Length 119;  
Best Local Similarity 69.7%; Pred. No. 1.1e-28;  
Matches 85; Conservative 11; Mismatches 21; Indels 5; Gaps 3;

OY 1 QVQLDSSGSGLVQAGSLRLSCAASGSIFFSDLMGWYRQAPGKEREAVALRT-RGTTSTY 59  
Db 1 EVQLVESGGGLVQPGGSLRLSCAASGFTFSYAMSWVRQAPGKLEWVSALISGSGSTTY 60  
OY 60 ADSYKGRFTISRDAKNTMYLQMNLSLKPEDTAVYYC-NARRSNYDRSGDYWGQGTQVTV 118  
Db 61 ADSYKGRFTISRDNKNTLYLQMNLSLRADTAVYYCAKDPGASY---YFDYWGQGTLYTV 117  
OY 119 SS 120  
Db 118 SS 119

## RESULT 15

S46391  
Ig heavy chain V region - human  
C/Species: Homo sapiens (man)  
C/Date: 27-Jan-1995 #sequence\_revision 27-Jan-1995 #text\_change 20-Jun-2000  
C/Accession: S46391  
R/Flgint, M./ Marks, J.D.; Winter, G.; Griffiths, A.D.  
J. Mol. Biol. 239, 68-78, 1994  
A/Title: In vitro assembly of repertoires of antibody chains on the surface of phage by  
A/Reference number: S46390; MUID:94254092; PMID:8196048  
A/Accession: S46391  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-114 <FIG>  
A/Cross-references: EMBL:Z31687; NID:9509784; PIDN:CAA83492.1; PID:91335144  
C:Superfamily: immunoglobulin V region; immunoglobulin homology  
C:Keywords: heterotetramer; immunoglobulin  
F:15-98/Domain: immunoglobulin homology <IMM>

Query Match 58.5%; Score 398; DB 2; Length 114;  
Best Local Similarity 69.4%; Pred. No. 1.1e-28;  
Matches 84; Conservative 8; Mismatches 21; Indels 8; Gaps 2;

OY 1 QVQLDSSGSGLVQAGSLRLSCAASGSIFFSDLMGWYRQAPGKEREAVALRT-RGTTSTY 59  
Db 1 QVNLRESGGGLVQPGGSLRLSCASGFTFSYAMHWYRQAPGKLEWVAIVYDGSNXY 60  
OY 60 ADSYKGRFTISRDAKNTMYLQMNLSLKPEDTAVYYCNARRSNYDRSGDYWGQGTQVTV 119  
Db 61 ADSYKGRFTISRDNKNTLYLQMNLSLRADTAVYYC-----ARDPGDYWGQGTLYTV 118  
OY 120 S 120  
Db 114 S 114

Search completed: October 5, 2004, 08:17:13  
Job time : 21.9028 secs

3 Page Blank (uspio)

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2004, 07:29:35 ; Search time 11.5463 Seconds  
(without alignments)  
561.749 Million cell updates/sec

Title: US-09-805-290A-18

Perfect score: 1 QVQLQDSGGGLVQAGGSLRL.....WGCGTQVTVSSAHSHSDPPS 129

Sequence: BLOSUM62

Scoring table: Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	373	54.9	136 1 HV16_MOUSE	P01783 mus musculu
2	366.5	53.9	115 1 HV3F_HUMAN	P01767 homo sapien
3	364	53.5	126 1 HV3K_HUMAN	P01772 homo sapien
4	363	53.4	114 1 HV3B_HUMAN	P01763 homo sapien
5	363	53.4	116 1 HV3T_HUMAN	P01781 homo sapien
6	356	52.4	120 1 HV3U_HUMAN	P01782 homo sapien
7	355	52.2	122 1 HV3G_HUMAN	P01768 homo sapien
8	350	51.5	116 1 HV05_CARP	P19181 carassius a
9	349.5	51.4	117 1 HV3C_HUMAN	P01764 homo sapien
10	344	50.6	122 1 HV3A_HUMAN	P01766 homo sapien
11	343.5	50.5	117 1 HV02_CANFA	P01785 canis famli
12	342.5	50.4	121 1 HV3J_HUMAN	P01771 homo sapien
13	341.5	50.2	115 1 HV3D_HUMAN	P01765 homo sapien
14	340.5	50.1	119 1 HV3I_HUMAN	P01770 homo sapien
15	339	49.9	114 1 HV01_CANFA	P01784 canis famli
16	335.5	49.3	120 1 HV3E_HUMAN	P01765 homo sapien
17	334.5	49.2	142 1 HV01_RAT	P01805 ratius norv
18	333.5	49.0	115 1 HV33_MOUSE	P01803 mus musculu
19	333	49.0	122 1 HV20_MOUSE	P01789 mus musculu
20	332.5	48.9	118 1 HV39_MOUSE	P01809 mus musculu
21	332	48.8	122 1 HV3H_HUMAN	P01769 homo sapien
22	331	48.7	119 1 HV3M_HUMAN	P01774 homo sapien
23	330.5	48.6	119 1 HV3L_HUMAN	P01773 homo sapien
24	330	48.5	117 1 HV3O_HUMAN	P01776 homo sapien
25	330	48.5	119 1 HV40_MOUSE	P01810 mus musculu
26	329.5	48.5	113 1 HV30_MOUSE	P01799 mus musculu
27	329.5	48.5	115 1 HV32_MOUSE	P01801 mus musculu
28	329	48.4	119 1 HV37_MOUSE	P01807 mus musculu
29	329	48.4	119 1 HV3N_HUMAN	P01775 homo sapien
30	327	48.1	97 1 HV56_MOUSE	P18527 mus musculu
31	327	48.1	111 1 HV35_MOUSE	P01804 mus musculu
32	326.5	48.0	117 1 HV53_MOUSE	P18524 mus musculu
33	325.5	47.9	117 1 HV55_MOUSE	P18526 mus musculu

34	325.5	47.9	123 1 HV22_MOUSE	P01791 mus musculu
35	325.5	47.9	123 1 HV25_MOUSE	P01794 mus musculu
36	325	47.8	119 1 HV38_MOUSE	P01808 mus musculu
37	324.5	47.7	113 1 HV27_MOUSE	P01796 mus musculu
38	324.5	47.7	117 1 HV54_MOUSE	P18525 mus musculu
39	323	47.5	119 1 HV3P_HUMAN	P01777 homo sapien
40	321.5	47.3	123 1 HV23_MOUSE	P01792 mus musculu
41	320.5	47.1	113 1 HV31_MOUSE	P01800 mus musculu
42	320.5	47.1	123 1 HV18_MOUSE	P01787 mus musculu
43	320.5	47.1	123 1 HV19_MOUSE	P01788 mus musculu
44	320.5	47.1	123 1 HV24_MOUSE	P01793 mus musculu
45	320	47.1	122 1 HV21_MOUSE	P01790 mus musculu

## ALIGNMENTS

RESULT 1	ID	Sequence	STANDARD	PRT	136 AA
AC	P01783	21-UTL-1986 (Rel. 01, Created)			
DT	21-UTL-1986 (Rel. 01, Last sequence update)				
DT	10-OCT-2003 (Rel. 42, Last annotation update)				
DE	Ig heavy chain V region MOPC 21 precursor (Fragment)				
OS	Mus musculus (Mouse)				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
OX	NCBI_TaxID=10090;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=8124548; PubMed=6788376;				
RA	Bothwell A.L.M., Paskind M., Rath M., Imanishi-Kari T., Rajewsky K.,				
RA	Baltimore D.;				
RT	"Heavy chain variable region contribution to the NFB family of				
RT	antibodies: somatic mutation evident in a gamma 2a variable region."				
RL	Cell 24:625-637 (1981).				
RN	[2]				
RP	SEQUENCE OF 17-136.				
RX	MEDLINE=77100368; PubMed=401950;				
RA	Adetunbo K., Milstien C., Secher D.S.;				
RL	"Molecular analysis of spontaneous somatic mutants."				
RL	Nature 263:299-304 (1977).				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>				
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).				
CC	-----				
DR	EMBL; J00522; AAD15290.1; -				
DR	PIR; E90809; GIMS21.				
DR	PDB; 1IGC; 03-JUN-95.				
DR	InterPro; IPR007110; Ig-like.				
DR	InterPro; IPR003596; Ig_V.				
DR	SMART; PF00047; IgV_1.				
DR	PROSITE; PS50835; IG_Like; 1.				
KW	Immunoglobulin V region; Signal; 3D-structure.				
FT	NON_TER	1			
FT	SIGNAL	<1	16		
FT	CHAIN	17	136		
FT	DOMAIN	115	119		
FT	DOMAIN	120	136		
FT	DISULFID	38	112		
FT	CONFLICT	75	78		
FT	CONFLICT	89	90		
FT	CONFLICT	115	115		
FT	CONFLICT	120	120		
FT	NON_TER	136	136		
SQ	SEQUENCE	136 AA;	15071 MW;	2276A98DBDBF7016 CRC64;	

```

Query Match.          54.9%; Score 373; DB 1; Length 136;
Best Local Similarity 67.5%; Pred. No. 8,8e-31;
Matches 81; Conservative 9; Mismatches 28; Indels 2; Gaps 2;

QY 2 VQLQDSGGGLVQAGSLRLSCAASGIFSSDLMGWYRQAPGKERAVARITRGTT-SYA 60
DB 18 VOLVESGGGLVQPGGSRKLSCAASGTFSSFGMHVYRQAPGKGLWVAIVISSGSTLHYA 77
QY 61 DSVKGFRTSRDNKATMYLQMSLKPEPTAVYYCNARSNYDRSGMDYWGQGTQVTVSS 120
DB 78 DTVKGFRTSRDNPKNITLFLQWTLRSSEDTANYTC-ARNGNVPYAMQYWGQGTSTVTVSS 136

RESULT 2
HV3F HUMAN          STANDARD; PRT; 115 AA.
AC P01767;
DT 21-UTL-1986 (Rel. 01, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ig heavy chain V-III region BUI.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
CX NCBI_Taxid=9606;
OC NBI_Taxid=9606;
RN [1]
RX MEDLINE=78137069; PubMed=416441;
RA Torano A., Putnam F.W.;
RT "Complete amino acid sequence of the alpha 2 heavy chain of a human
RT IgA2 immunoglobulin of the A2m (2) allotype."
CC CC -1- MISCELLANEOUS: THE SEQUENCE OF THE ALPHA-2, A2M(2) ALLOTYPIC, C
CC CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.
CC CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.
DR HSP; P01789; IMCP.
DR GO; GO:0005576; C:extracellular; NAS.
DR GO; GO:0003823; F:antigen binding; NAS.
DR GO; GO:0006955; P:immune response; NAS.
DR InterPro; IPR007110; Ig_V.
DR InterPro; IPR003596; Ig_V.
DR Pfam; PF00047; Ig_1.
DR SMART; SM00406; IGV_1.
DR PROSITE; PS00835; IG_LIKE; 1.
KM Immunoglobulin V region.
FT DOMAIN 1 111 IG-LIKE.
FT NON_TER 115 115
SQ SEQUENCE 115 AA; 12379 MW; 208876A7DF52DCF4 CRC64;

Query Match          53.9%; Score 366.5; DB 1; Length 115;
Best Local Similarity 63.3%; Pred. No. 3,3e-30;
Matches 76; Conservative 16; Mismatches 23; Indels 5; Gaps 2;

QY 1 QVQLDSSGGLVQAGSLRLSCAASGIFSSDLMGWYRQAPGKERAVARITRGTT-SYA 60
DB 1 EVQLVETGGGLIQPGGSRKLSCAASGTFVBSHSMVYRQAPGKALZWVSAITRGTTIYA 60
QY 61 DSVKGFRTSRDNKATMYLQMSLKPEPTAVYYCNARSNYDRSGMDYWGQGTQVTVSS 120
DB 61 DSVKGFRTSRDSDSRRTVTLQWESLRAEDTAVYYC-AR---DLAALFLFGKGTIVTVSS 115

RESULT 3
HV3F HUMAN          STANDARD; PRT; 126 AA.
AC P01772;
DT 21-UTL-1986 (Rel. 01, Created)
DT 21-UTL-1986 (Rel. 01, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ig heavy chain V-III region KOL.
OS Homo sapiens (Human).

```

```

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
CX NCBI_Taxid=9606;
RN [1]
RX MEDLINE=83289131; PubMed=6884994;
RA Schmidt W.E., Jung H.-D., Palm W., Hilschmann N.;
RT "Three-dimensional structure determination of antibodies. Primary
RT structure of crystallized monoclonal immunoglobulin IgG1 KOL, I."
RL Hoppe-Seyler's Z. Physiol. Chem. 364:713-747(1983).
RN [2]
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).
RX MEDLINE=81072235; PubMed=7441755;
RA Marquart M., Detschenhofer J., Huber R., Palm W.;
RT "Crystallographic refinement and atomic models of the intact
RT immunoglobulin molecule KOL and its antigen-binding fragment at 3.0 A
RT and 1.0-A resolution."
RL J. Mol. Biol. 141:369-391(1980).
CC CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.
CC CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.
DR PIR; A02055; G1HUKL.
DR PDB; 2FB4; 12-UTL-89.
DR GO; GO:0005576; C:extracellular; NAS.
DR GO; GO:0003823; F:antigen binding; NAS.
DR GO; GO:0006955; P:immune response; NAS.
DR InterPro; IPR007110; Ig_Like.
DR InterPro; IPR003596; Ig_V.
DR Pfam; PF00047; Ig_1.
DR SMART; SM00406; IGV_1.
DR PROSITE; PS00835; IG_LIKE; 1.
KM Immunoglobulin V region; 3D-structure; Pyrrolidone carboxylic acid.
FT DOMAIN 1 112 IG-LIKE.
FT DISULFID 22 96 PYRROLIDONE CARBOXYLIC ACID.
FT DISULFID 105 110
FT STRAND 3 7
FT STRAND 11 12
FT STRAND 14 15
FT STRAND 18 25
FT HELIX 29 31
FT STRAND 34 39
FT STRAND 41 42
FT STRAND 45 51
FT STRAND 53 54
FT STRAND 58 60
FT HELIX 62 64
FT STRAND 65 65
FT STRAND 66 67
FT STRAND 68 72
FT TURN 74 77
FT STRAND 78 83
FT HELIX 88 90
FT STRAND 92 99
FT STRAND 106 106
FT TURN 107 108
FT STRAND 109 109
FT STRAND 113 116
FT STRAND 120 124
FT NON_TER 126 126
SQ SEQUENCE 126 AA; 13718 MW; E4D71B52B16F8776 CRC64;

Query Match          53.5%; Score 364; DB 1; Length 126;
Best Local Similarity 63.8%; Pred. No. 6,5e-30;
Matches 81; Conservative 13; Mismatches 25; Indels 8; Gaps 4;

QY 1 QVQLDSSGGLVQAGSLRLSCAASGIFSSDLMGWYRQAPGKERAVARITRGTT-SY 59
DB 1 QVQLVSGGGLVQPGGSRKLSLSCSSSGFRTSSVAMVYRQAPGKGLWVAIINDGSDQHY 60
QY 60 ADSVKGFRFTSRDNKATMYLQMSLKPEPTAVYYCNARSNYD-----RSMG-DYWGQG 113
DB 61 ADSVKGFRFTSRDSDSRRTVTLQWDSLRPEDTGYFC-ARDGGHGFCCSSASCSPDYWGQ 119

```

QY 114 TQTVSS 120  
 DB 120 TQTVSS 126

## RESULT 4

HV3B\_HUMAN  
 ID HV3B\_HUMAN STANDARD; PRT; 114 AA.  
 AC P01763;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Ig heavy chain V-III region WEA.  
 OS Homo sapiens (Human)  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 NCBI\_TaxID=9606;  
 RN [1]  
 RP MEDLINE=83273707; PubMed=6410398;  
 RX GO:0005576; C:extracellular; NAS.  
 RA "Amino acid sequence of the Fv region of a human monoclonal IgM (protein WEA) with antibody activity against 3,4-pyruvylated galactose in Klebsiella polysaccharides K30 and K33."  
 RT Proc. Natl. Acad. Sci. U.S.A. 80:4837-4841 (1983).  
 DR InterPro: IPR007110; IG-like.  
 DR InterPro: IPR003596; IG\_v.  
 DR Pfam: PF00047; IG; 1.  
 DR SMART: SM00406; IGv; 1.  
 DR PROSITE: PS50835; IG-like; 1.  
 KM Immunoglobulin V region; Pyroglutamate carboxylic acid.  
 FT DOVAIN 1 112  
 FT MOD\_RES 1 112  
 FT NON\_TER 1 114  
 FT SEQUENCE 114 AA; 12256 MW; D88294FB418A07B7 CRC64;  
 Query Match 53.4%; Score 363; DB 1; Length 114;  
 Best Local Similarity 63.6%; Pred. No. 7.3e-30;  
 Matches 77; Conservative 14; Mismatches 22; Indels 8; Gaps 2;

OX NCBI\_TaxID=9606;  
 RN [1]  
 RP MEDLINE=75059123; PubMed=4803843;  
 RX GO:0005576; C:extracellular; NAS.  
 RA Watanabe S., Barnikol H.U., Horn J., Bertram J., Hilschmann N.,  
 RT "The primary structure of a monoclonal IgM-immunoglobulin (macroglobulin Gal.), II: the amino acid sequence of the H-chain (mu-type), subgroup H III. Architecture of the complete IgM-molecule."  
 RT Hoppe-Seyler's Z. Physiol. Chem. 354:1505-1509 (1973).  
 RN [2]  
 RP REVISION TO 28-33.  
 RA Hilschmann N.,  
 RA Submitted (JUN-1975) to the PIR data bank.  
 CC -1- MISCELLANEOUS: THIS MU CHAIN WAS ISOLATED FROM A WALDENSTROM'S MACROGLOBULIN.  
 CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.  
 DR HSSP: P01772; 2FB4.  
 DR GO:0005576; C:extracellular; NAS.  
 DR GO:0003823; F:antigen binding; NAS.  
 DR GO:0006955; P:immune response; NAS.  
 DR InterPro: IPR007110; IG-like.  
 DR InterPro: IPR003596; IG\_v.  
 DR Pfam: PF00047; IG; 1.  
 DR SMART: SM00406; IGv; 1.  
 DR PROSITE: PS50835; IG-like; 1.  
 KM Immunoglobulin V region.  
 FT DOVAIN 1 112  
 FT MOD\_RES 1 112  
 FT NON\_TER 1 116  
 FT SEQUENCE 116 AA; 12730 MW; 2C67CA9AAAAA1282 CRC64;  
 Query Match 53.4%; Score 363; DB 1; Length 116;  
 Best Local Similarity 64.2%; Pred. No. 7.5e-30;  
 Matches 79; Conservative 8; Mismatches 26; Indels 10; Gaps 3;

QY 1 QVQLDSGGGLVQAGSLRLSCAASGIFSSDLMGWYRQAPKEREAVARITRGST-59  
 DB 1 QVQLVDSGGGLVQAGSLRLSCAASGIFSSDLMGWYRQAPKEREAVARITRGST-59  
 QY 60 ADSYKGRFTISPDANKMTWYLNQNSLKPEDTAVYYCNARSRNSYDRSW--GDYWGQGTQV 117  
 DB 61 ADSYKGRFTISPDANKMTWYLNQNSLKPEDTAVYYC-----ARGWLNWGGGLTAVTS 113

QY 1 QVQLDSGGGLVQAGSLRLSCAASGIFSSDLMGWYRQAPKEREAVARITRGST-59  
 DB 1 EQVLVDSGGGLVQAGSLRLSCAASGIFSSDLMGWYRQAPKEREAVARITRGST-59  
 QY 60 ADSYKGRFTISPDANKMTWYLNQNSLKPEDTAVYYCNARSRNSYDRSW--GDYWGQGTQV 117  
 DB 61 VDSYKGRFTISPDANKMTWYLNQNSLKPEDTAVYYC-----ARGWLNWGGGLTAVTS 113

RESULT 6  
 HV3U\_HUMAN  
 ID HV3U\_HUMAN STANDARD; PRT; 120 AA.  
 AC P01782;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Ig heavy chain V-III region DOB.  
 OS Homo sapiens (Human)  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 NCBI\_TaxID=9606;  
 RN [1]  
 RP MEDLINE=80020921; PubMed=114209;  
 RX Steiner L.A., Garcia Pardo A., Margolies M.N.,  
 RT "Amino acid sequence of the heavy-chain variable region of the crystallizable human myeloma protein Dob."  
 RT Biochemistry 18:4068-4080 (1979).  
 RN [2]  
 RP CRYSTALLIZATION.  
 RX MEDLINE=80020920; PubMed=114208;  
 RX Steiner L.A., Lopes A.D.,  
 RT "The crystallizable human myeloma protein Dob has a hinge-region deletion."





```

ID HV3C_HUMAN STANDARD; PRT; 117 AA.
AC P01764;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DI 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ig heavy chain V-II region VH26 precursor.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=81101090; PubMed=6450418;
RA Mathysens G., Rabbits T.H.;
RT "Structure and multiplicity of genes for the human immunoglobulin
RT heavy chain variable region."
RL Proc. Natl. Acad. Sci. U.S.A. 77:6561-6565(1980).
CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; J00236; AAAS3516.1; -
DR EMBL; M35415; AAAS8735.1; -
DR PIR; A02047; H3HUC26.
DR PDB; 1HOU; 23-DEC-99.
DR Genew; HGNC:5545; IGHV@
DR GO; GO:0005576; C:extracellular; NAS.
DR GO; GO:0003823; F:antigen binding; NAS.
DR GO; GO:0006955; P:immune response; NAS.
DR InterPro; IPR007110; IG-1-like.
DR InterPro; IPR003596; IG_V.
DR Pfam; PF00047; Ig_1.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS50835; IG_LIKE; 1.
DR Immunoglobulin V region; Signal; 3D-structure.
KW CHAIN 1
FT SIGNAL 19
FT CHAIN 20 117 IG HEAVY CHAIN V-II REGION VH26.
FT DOMAIN 20 >117 IG-LIKE.
FT NON_TER 117
FT SEQUENCE 117 AA; 12582 MW; E826733F1A3CB0F1 CRC64;

Query Match 51.4%; Score 349.5; DB 1; Length 117;
Best Local Similarity 74.0%; Pred. No. 1.7e-28;
Matches 71; Conservative 8; Mismatches 16; Indels 1; Gaps 1;

QY 1 QVQLDSGGGLVQAGSLRLSCAASGSIFFSDLMGWRQAPGKERAVART-RGRTSY 59
DB 20 EVQLLESGGGLVQPGGSRRLSCASGFTFSYAMSWRQAKGLEWVSALISGSGSTYY 79
QY 60 ADSVKGRTISRDNKATMYLQWNSLKPEDTAVYYC 95
DB 80 GDSVKGRTISRDNKATMYLQWNSLKPEDTAVYYC 115

```

```

RP SEQUENCE (MYELOMA PROTEIN TRO).
RX MEDLINE=76023781; PubMed=809331;
RA Kratzin H., Altevogt P., Ruben E., Kortt A., Starosiek K.,
RA Hilschmann N.;
RT "The primary structure of a monoclonal IgA-immunoglobulin (IgA Tro.),
RT II. The amino acid sequence of the H-chain, alpha-type, subgroup III;
RT structure of the complete IgA-molecule."
RL Hoppe-Seyler's Z. Physiol. Chem. 356:1337-1342(1975).
CC -1- MISCELLANEOUS: THE SEQUENCE OF THE C REGION IS ALSO GIVEN.
CC -----
CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.
CC -----
DR PIR; A02045; ALHUR.
DR HSSP; P01772; 2FB4.
DR GO; GO:0005576; C:extracellular; NAS.
DR GO; GO:0003823; F:antigen binding; NAS.
DR GO; GO:0006955; P:immune response; NAS.
DR InterPro; IPR007110; IG-1-like.
DR InterPro; IPR003596; IG_V.
DR Pfam; PF00047; Ig_1.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS50835; IG_LIKE; 1.
DR Immunoglobulin V region; Pyrolydine carboxylic acid.
FT DOMAIN 1 108
FT MOD_RES 1 108 PYRROLIDONE CARBOXYLIC ACID.
FT NON_TER 122 122
FT SEQUENCE 122 AA; 13472 MW; 2E21A11DA04D80F9 CRC64;

Query Match 50.6%; Score 344; DB 1; Length 122;
Best Local Similarity 56.0%; Pred. No. 6.6e-28;
Matches 70; Conservative 18; Mismatches 29; Indels 8; Gaps 4;

QY 1 QVQLDSGGGLVQAGSLRLSCAASGSIFFSDLMGWRQAPGKERAVART-RGRTSY 59
DB 1 QVQLVDSGGGLVQPGGSRRLSCVASFSPRDEPMVIRTPGGLTWSYVIGSGSGTLY 60
QY 60 ADSVKGRTISRDNKATMYLQWNSLKPEDTAVYYCNARSDRNG----DYWGCGTG 115
DB 61 ADSVKGRTISRDNKATMYLQWNSLKPEDTAVYYC-AATBEB--WTFSLRYWCGZGL 117
QY 116 VTWSS 120
DB 118 VTWSS 122

RESULT 11
ID HV02_CANFA STANDARD; PRT; 117 AA.
AC P01785;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DI 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ig heavy chain V region MOO.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE OF 1-112.
RX MEDLINE=77242268; PubMed=407924;
RA Wasserman R.L.; Capra J.D.;
RT "Primary structure of the variable regions of two canine
RT immunoglobulin heavy chains."
RL Biochemistry 16:3160-3168(1977).
RN [2]
RP SEQUENCE OF 113-117.
RX MEDLINE=80077682; PubMed=117299;
RA McCumber L.V.; Capra J.D.;
RT "The complete amino-acid sequence of a canine mu chain."
RL Mol. Immunol. 16:565-570(1979).
CC -1- MISCELLANEOUS: THIS MU CHAIN WAS ISOLATED FROM A MYELOMA PROTEIN.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.
DR PIR; A04003; MEDGO.
DR HSSP; P01772; 2FB4.
DR InterPro; IPR007110; IG-1-like.

```

DR InterPro: IPR003596; IG\_V.  
 DR Pfam: PF00047; IG\_1.  
 DR SMART: SM00406; IGV; 1.  
 DR PROSITE: PS50835; IG-LIKE; 1.  
 KW Immunoglobulin V region.  
 FT DOMAIN 1 116 IG-LIKE.  
 FT NON TER 117 117  
 SQ SEQUENCE 117 AA; 12703 MW; FCE07309E0A84B35 CRC64;

Query Match 50.5%; Score 343.5; DB 1; Length 117;  
 Best Local Similarity 59.2%; Pred. No. 7e-28;  
 Matches 74; Conservative 13; Mismatches 25; Indels 13; Gaps 2;

QY 1 QVQLDSSGGGLVQAGSLRLSCAASGSIFFSPDLMGVYRQAPGKEREAVARITRGSTTSYA 60  
 DB 1 EVKLVESSGGDLVPRGSLRLSCVAGSFTFSSNGMSWVRQDPGEGLQWADISSSGQRTYA 60  
 QY 61 DSVKGRFTISRDNANKNTYLMQNSLKPEDTAVYYCNARRSYNDRSWGDP-----YWGQCTQ 115  
 DB 61 DAVKGRFTISRDNANKNTYLYLQMEDLRVEDTAVYYCATE-----GDIEIPRYGQGTI 112

QY 116 VTWSS 120  
 DB 113 VTWSS 117

RESULT 12  
 HVJ3 HUMAN STANDARD; PRT; 121 AA.

AC P01771; 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Ig heavy chain V-II region HIL.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RX MEDLINE=79124695; PubMed=420800;  
 RA Chiu Y.-Y.H., Lopez de Castro J.A., Poljak R.J.;  
 RT "Amino acid sequence of the VH region of human myeloma  
 RL cytimmunoglobulin Igg H11";  
 CC Biochemistry 18:353-360(1979).  
 CC -1- MISCELLANEOUS: THIS CHAIN WAS ISOLATED FROM AN IGG1 MYELOMA  
 CC PROTEIN.

CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.  
 DR PIR; A02054; GIHUL.  
 DR HSSP; P01772; 2FB4.  
 DR GO; GO:0005576; C:extracellular; NAS.  
 DR GO; GO:0003823; P:antigen binding; NAS.  
 DR GO; GO:0006955; P:immune response; NAS.  
 DR InterPro: IPR007110; IG-Like.  
 DR InterPro: IPR003596; IG\_V.  
 DR Pfam: PF00047; IG\_1.  
 DR SMART: SM00406; IGV; 1.  
 DR PROSITE: PS50835; IG-LIKE; 1.  
 KW Immunoglobulin V region; Pyridone carboxylic acid.  
 FT DOMAIN 1 112 IG-LIKE.  
 FT MOD RES 1 1 PYRROLIDONE CARBOXYLIC ACID.  
 FT NON TER 121 121  
 SQ SEQUENCE 121 AA; 13566 MW; 480FC53610E5DAB CRC64;

Query Match 50.4%; Score 342.5; DB 1; Length 121;  
 Best Local Similarity 62.3%; Pred. No. 9.2e-28;  
 Matches 76; Conservative 12; Mismatches 31; Indels 3; Gaps 3;

QY 1 QVQLDSSGGGLVQAGSLRLSCAASGSIFFSPDLMGVYRQAPGKEREAVARITRGSTTSY 59  
 DB 1 QVQLVAGAGGVQPGSLRLSCVAGSFTFSSNGMSWVRQDPGEGLQWADISSSGQRTYA 60  
 QY 60 ADSVKGKFTISRDNANKNTYLMQNSLKPEDTAVYYCNARRSYNDRSWG-DYWGQGTQTV 118

DB 61 GDSVKGRFTISRDNANKNTYLMQNSLKPEDTAVYYC-ARPDILTAFSFDYWGQGLVTV 119  
 QY 119 SS 120  
 DB 120 SS 121

RESULT 13  
 HVJ3 HUMAN STANDARD; PRT; 115 AA.

AC P01765; 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Ig heavy chain V-II region TIL.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RX MEDLINE=78005528; PubMed=409716;  
 RA Wang A.-C., Wang I.-Y., Fudenberg H.H.;  
 RT "Immunoglobulin structure and genetics. Identity between variable  
 RT regions of a mu and a gamma2 chain";  
 RL J. Biol. Chem. 252:7192-7199(1977).  
 CC -1- MISCELLANEOUS: THE SEQUENCES OF THE V REGIONS OF THE HEAVY CHAINS  
 CC OF IGM AND IGG2 ISOLATED FROM A SINGLE PATIENT WITH BICLONAL  
 CC GAMOPATHY ARE IDENTICAL. THEIR LIGHT CHAINS ARE APPARENTLY ALSO  
 CC IDENTICAL.

CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.  
 DR PIR; A02048; H3HUL.  
 DR HSSP; P01772; 2FB4.  
 DR GO; GO:0005576; C:extracellular; NAS.  
 DR GO; GO:0003823; P:antigen binding; NAS.  
 DR GO; GO:0006955; P:immune response; NAS.  
 DR InterPro: IPR007110; IG-Like.  
 DR InterPro: IPR003596; IG\_V.  
 DR Pfam: PF00047; IG\_1.  
 DR SMART: SM00406; IGV; 1.  
 DR PROSITE: PS50835; IG-LIKE; 1.  
 KW Immunoglobulin V region.  
 FT DOMAIN 1 108 IG-LIKE.  
 FT NON TER 115 115  
 SQ SEQUENCE 115 AA; 12356 MW; 4DC67D179F62326 CRC64;

Query Match 50.2%; Score 341.5; DB 1; Length 115;  
 Best Local Similarity 60.3%; Pred. No. 1.1e-27;  
 Matches 73; Conservative 14; Mismatches 27; Indels 7; Gaps 3;

QY 1 QVQLDSSGGGLVQAGSLRLSCAASGSIFFSPDLMGVYRQAPGKEREAVARITRGSTTSY 59  
 DB 1 EVQLDSGGGLVQPGSLRLSCAASGFTSTYMSWVRQDPGGLZVGAIGLSYSZSY 60  
 QY 60 ADSVKGKFTISRDNANKNTYLMQNSLKPEDTAVYYCNARRSYNDRSWGQYWGQGTQTV 119  
 DB 61 ADSVKGKFTISRDNANKNTYLMQNSLKPEDTAVYYC-AKGVSAVYFPMWQZGLTVTV 114

QY 120 S 120  
 DB 115 S 115

RESULT 14  
 HVJ3 HUMAN STANDARD; PRT; 119 AA.

AC P01770; 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Ig heavy chain V-II region NIE.  
 OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_TaxId=9606;  
RN [1]  
RP MEDLINE=77070269; PubMed=826475;  
RA Postnigl H., Hilschmann N.;  
RT "The rule of antibody structure. The primary structure of a  
RT monoclonal IgG1 immunoglobulin (myeloma protein Nle). II. The  
RT chymotryptic peptides of the H-chain, alignment of the tryptic  
RT peptides and discussion of the complete structure."  
RL Hoppe-Seyler's Z. Physiol. Chem. 357:1571-1604 (1976).  
RN [2]  
RP DISULFIDE BOND.  
RX MEDLINE=77070267; PubMed=1002129;  
RA Dreker L., Schwarz U., Reiche W., Hilschmann N.;  
RT "Rule of antibody structure. The primary structure of a monoclonal  
RT IgG1 immunoglobulin (myeloma protein Nle). I: Purification and  
RT characterization of the protein, the L- and H-chains, the  
RT cyanogen bromide cleavage products, and the disulfide bridges."  
RL Hoppe-Seyler's Z. Physiol. Chem. 357:1515-1540 (1976).  
CC -1- MISCELLANEOUS: THIS CHAIN WAS ISOLATED FROM AN IG1 MYELOMA  
CC PROTEIN.  
CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.  
DR PIR; A91668; GIHUNI.  
DR HSSP; P01772; 2FB4.  
DR GO; GO:0005576; C:extracellular; NAS.  
DR GO; GO:0003823; F:antigen binding; NAS.  
DR GO; GO:0006955; P:immune response; NAS.  
DR InterPro; IPR007110; Ig-like.  
DR InterPro; IPR003596; Ig\_v.  
DR Pfam; PF00047; Ig; 1.  
DR SMART; SM00406; IGV; 1.  
DR PROSITE; PS50835; IG LIKE; 1.  
KW Immunoglobulin V region; Pyrolydione carboxylic acid.  
FT DOMAIN 1 112 IG-LIKE.  
FT MOD\_RSS 1 1 PYROLYDIONE CARBOXYLIC ACID.  
FT DISULFID 22 96  
FT NON\_TER 119 119  
SQ SEQUENCE 119 AA; 13242 MW; C96935A655E5B5B CRC64;  
Query Match 50.1%; Score 340.5; DB 1; Length 119;  
Best Local Similarity 62.0%; Pred. No. 1.4e-27;  
Matches 75; Conservative 11; Mismatches 32; Indels 3; Gaps 2;  
Db 1 QVQLDGGGGLVQAGSLRLSCAASGSIFFSDLMGWYRQAPKEREAVARIT-RCGTTSTY 59  
1 QVQLVQSGGGLVQAGSLRLSCAASGFTFSRYTHHWRAQAPKGLQVAVYMSYBGBKHV 60  
QY 60 ADSVKGFTISRDNKNTMYLQMSLKPEPTAVYYCNARNSNYDRSKG-DYWGQGTQVTV 119  
Db 61 ADSVNGRFTISRDNKNTLYLQMSLKPEPTAVYYCARIRDT--AMFFAHWGQGTQVTV 118  
QY 120 S 120  
Db 119 S 119  
RESULT 15  
HVO1\_CANFA STANDARD; PRT; 114 AA.  
AC P01784;  
DT 21-JUL-1986 (Ref. 01, Created)  
DT 21-JUL-1986 (Ref. 01, Last sequence update)  
DT 10-OCT-2003 (Ref. 42, Last annotation update)  
DE 19 heavy chain V region GOM.  
OS Canis familiaris (Dog).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.  
OX NCBI\_TaxId=9615;  
RN [1]  
RP SEQUENCE.  
RX MEDLINE=77242268; PubMed=407924;

RA Wasserman R.L., Capra J.D.;  
RT "Primary structure of the variable regions of two canine  
RT immunoglobulin heavy chains."  
RL Biochemistry 16:3160-3168 (1977).  
CC -1- MISCELLANEOUS: THIS CHAIN WAS ISOLATED FROM A MYELOMA PROTEIN.  
CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.  
DR PIR; A02067; AYDCGM.  
DR HSSP; P01772; 2FB4.  
DR InterPro; IPR007110; Ig-like.  
DR InterPro; IPR003596; Ig\_v.  
DR Pfam; PF00047; Ig; 1.  
DR SMART; SM00406; IGV; 1.  
DR PROSITE; PS50835; IG LIKE; 1.  
KW Immunoglobulin V region.  
FT DOMAIN 1 112 IG-LIKE.  
FT NON\_TER 114 114  
SQ SEQUENCE 114 AA; 12430 MW; B1D4745D2C4E13C4 CRC64;  
Query Match 49.9%; Score 339; DB 1; Length 114;  
Best Local Similarity 59.8%; Pred. No. 1.9e-27;  
Matches 73; Conservative 11; Mismatches 28; Indels 10; Gaps 3;  
Db 1 QVQLDGGGGLVQAGSLRLSCAASGSIFFSDLMGWYRQAPKEREAVARITRG-GTTSY 59  
1 EVQLVESGGDLVQPGSLRLSCVASGITFSGYDMQWRQAPKGLQVAVYFNDALSAQGY 60  
QY 60 ADSVKGFTISRDNKNTMYLQMSLKPEPTAVYYCNARNSNYDRSKG-DYWGQGTQVTV 118  
Db 61 ADSVNGRFTISRDNKNTLYLQMSLKPEPTAVYYC-----APWQFETWGQGTQVTV 112  
QY 119 SS 120  
Db 113 SS 114  
Search completed: October 5, 2004, 08:01:53  
Job time: 11.5463 secs

This Page Blank (uspo10)

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2004, 07:34:11 ; Search time 68.2824 Seconds  
(without alignments)  
596.081 Million cell updates/sec

Title: US-09-805-290A-18

Perfect score: 1 QVQLDSSGGLVQAGGSLRL.....WGQGVTVSAHSEDPSS 129

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

SPREMBL.25:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phase:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriaph:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	400.5	58.9	597	4 Q96BB9	Q96BB9 homo sapien
2	399	58.7	118	4 Q9UL72	Q9UL72 homo sapien
3	379.5	55.8	119	11 Q92OE7	Q92OE7 mus musculu
4	379.5	55.8	131	4 Q9UL88	Q9UL88 homo sapien
5	376.5	55.4	113	4 Q9UL90	Q9UL90 homo sapien
6	375	55.1	613	4 Q8WUK1	Q8WUK1 homo sapien
7	373.5	54.9	147	4 Q9UL71	Q9UL71 homo sapien
8	373.5	54.9	147	4 Q9Y509	Q9Y509 homo sapien
9	371	54.6	112	4 Q9HCC1	Q9HCC1 homo sapien
10	369	54.3	118	4 Q9UL91	Q9UL91 homo sapien
11	367.5	54.0	499	4 Q8NSK4	Q8NSK4 mus musculu
12	367	54.0	487	11 Q99KA4	Q99KA4 mus musculu
13	364	53.5	573	4 Q8WUJ8	Q8WUJ8 mus musculu
14	361.5	53.2	116	4 Q9UL93	Q9UL93 mus musculu
15	360	52.9	473	11 Q91Z05	Q91Z05 mus musculu
16	360	52.9	494	4 Q96K68	Q96K68 homo sapien

17	353.5	52.0	486	11 Q91Z07	Q91Z07 mus musculu
18	353	51.9	122	4 Q9UL84	Q9UL84 homo sapien
19	350.5	51.5	480	11 Q91XE1	Q91XE1 mus musculu
20	350	51.5	298	11 Q9QYF0	Q9QYF0 mus musculu
21	348.5	51.2	479	11 Q91WP5	Q91WP5 mus musculu
22	345.5	50.8	493	4 Q8NCL6	Q8NCL6 mus musculu
23	344	50.6	469	11 Q8R3V9	Q8R3V9 mus musculu
24	342.5	50.4	95	4 Q9UL56	Q9UL56 mus musculu
25	341.5	50.2	487	11 Q80Z17	Q80Z17 mus musculu
26	337	49.6	437	11 Q9RIA4	Q9RIA4 mus musculu
27	335	49.3	521	4 Q8N4Y9	Q8N4Y9 mus musculu
28	332	48.8	470	4 Q7Z5M1	Q7Z5M1 mus musculu
29	318	46.8	484	11 Q8VBA0	Q8VBA0 mus musculu
30	310	45.6	124	6 Q9N0W6	Q9N0W6 cyctolaqus
31	308	45.3	124	6 Q9NDW4	Q9NDW4 cyctolaqus
32	303	44.6	479	11 Q7TKR4	Q7TKR4 mus musculu
33	302.5	44.5	112	4 Q9UGP3	Q9UGP3 mus musculu
34	297	43.7	482	4 Q7Z351	Q7Z351 mus musculu
35	291.5	42.9	104	4 Q9UL87	Q9UL87 mus musculu
36	287.5	42.8	482	11 Q91X92	Q91X92 mus musculu
37	287.5	42.3	473	11 Q9D8L4	Q9D8L4 mus musculu
38	283	41.6	145	11 Q924R4	Q924R4 mus musculu
39	280.5	41.2	119	4 Q9UL73	Q9UL73 mus musculu
40	280	41.2	124	4 Q9UL92	Q9UL92 mus musculu
41	279	41.0	145	11 Q924Q7	Q924Q7 mus musculu
42	276	40.6	482	11 Q8K172	Q8K172 mus musculu
43	275.5	40.5	614	11 Q7TMT6	Q7TMT6 mus musculu
44	274.5	40.4	146	11 Q924Q3	Q924Q3 mus musculu
45	271.5	39.9	614	4 Q96GA6	Q96GA6 mus musculu

ALIGNMENTS

RESULT 1		ID		Q96BB9	PRELIMINARY;	PRR;	597 AA.
AC	Q96BB9;						
DT	01-DEC-2001 (TREMBLrel. 19, Created)						
DT	01-DEC-2001 (TREMBLrel. 19, Last sequence update)						
DT	01-OCT-2003 (TREMBLrel. 25, Last annotation update)						
DE	Hypothetical protein.						
OS	Homo sapiens (Human).						
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;						
OC	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.						
OX	NCBI_TaxID=9606;						
RN	[1]						
RP	SEQUENCE FROM N.A.						
RC	TISSUE=B-cell;						
RA	Strasbourg R.;						
RL	Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.						
DR	EMBL: BC015760; AAH15760.1; -						
DR	InterPro; IPR007110; IG-like.						
DR	InterPro; IPR003006; IG_MHC.						
DR	InterPro; IPR003596; IG_V.						
DR	Pfam; PF00047; Ig; 5.						
DR	SMART; SM00406; IGV; 1.						
DR	PROSITE; PS50835; IG_LIKE; 5.						
DR	PROSITE; PS00290; IG_MHC; 3.						
KW	Hypothetical protein						
SQ	SEQUENCE 597 AA; 65039 MW; 4FCAD8BCE263D9 CRC64;						
Query Match		58.9%;	Score 400.5;	DB 4;	Length 597;		
Best Local Similarity		65.6%;	Pred. No. 1.5e-32;				
Matches 86;		Conservative 11;	Mismatches 25;	Indels 9;	Gaps 3;		
QY	1	QVQLDSSGGLVQAGGSLRLSCAASGIFSSDLMGWVROAPGKREAVARIT-NGGTSY	59				
DB	20	EVQLLESQGLVQPGSLRLSCAASGIFSSYANMWVROAPGKLEWVAISGGSGSY	79				
QY	60	ADSVKGRFTISRDNKNTMYLQNNLSLKPEDTAVYYC-----NARRSNYDRSGWDYGGG	113				
DB	80	ADSVKGRFTISRDNKNTMYLQNNLSLKPEDTAVYYCAKDPGYSASGNYTSE--DIWGCG	137				

QY 114 TQTVSSAHHS 124  
 Db 138 TLVTWSSGSSAS 148

## RESULT 2

Q9UL72 PRELIMINARY; PRT; 118 AA.

AC Q9UL72;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
 DE Myosin-reactive immunoglobulin heavy chain variable region  
 (Fragment).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98277139; PubMed=9614934;  
 RA Wu X., Liu B., Van der Werwe P.L., Kalis N.N., Berney S.M.,  
 RA Young D.C.;  
 RT "Myosin-reactive autoantibodies in rheumatic carditis and normal  
 fetus";  
 RL Clin. Immunol. Immunopathol. 87:184-192(1998).  
 DR EMBL; AF035042; AAD56278.1; -.  
 DR PIR; S21205; S21205.  
 DR HSSP; P01772; 2F84.  
 DR InterPro; IPR007110; Ig-like.  
 DR InterPro; IPR003596; Ig\_v.  
 DR Pfam; PF00047; Ig; 1.  
 DR SMART; SM00406; IGV; 1.  
 DR PROSITE; PS50835; IG\_LIKE; 1.  
 FT NON\_TER  
 FT NON\_TER  
 SQ SEQUENCE 118 AA; 12872 MW; B4D1A5944BD25CCA CRC64;

Query Match 58.7%; Score 399; DB 4; Length 118;  
 Best Local Similarity 67.5%; Pred. No. 2.6e-33;

Matches 81; Conservative 11; Mismatches 26; Indels 2; Gaps 1;

QY 1 QVQLDSSGGGLVQAGSLRLSCAASGTFSSDLMGWRQAPGKERAVARITRGCTSYA 60  
 Db 1 EVQLVESGGGLVQPGGSLRLSCAASGFTVSSNNMWVQAQGLKESVYLVSSGSSYYA 60  
 QY 61 DSVKGRFTISRDNKNTMYLQMNSLKEPEDTAVYYCCNARSNYDRSGWDYWGCGTQTVSS 120  
 Db 61 DSVKGRFTISRDNKNTMYLQMNSLKEPEDTAVYYC--ARDRFGFLPDYWGCGTLTVSS 118

## RESULT 3

Q920E7 PRELIMINARY; PRT; 119 AA.

AC Q920E7;  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
 DE Pterin-mimicking anti-Idiotope heavy chain variable region  
 (Fragment).  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Atkin J.D., Tape A., Jennings I.G., Horatits O., Cotton R.G.H.;  
 RT "Definition of the Idiotope of Pterin-Mimicking Antibodies Expressed  
 in Mamalian Cells";  
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF307937; AAL09421.1; -.  
 DR PIR; C25913; C25913.

DR InterPro; IPR007110; Ig-like.  
 DR InterPro; IPR003596; Ig\_v.  
 DR Pfam; PF00047; Ig; 1.  
 DR SMART; SM00406; IGV; 1.  
 DR PROSITE; PS50835; IG\_LIKE; 1.  
 FT NON\_TER  
 FT NON\_TER  
 SQ SEQUENCE 119 AA; 13025 MW; F6B904044381CA7C CRC64;

Query Match 55.8%; Score 379.5; DB 11; Length 119;  
 Best Local Similarity 66.9%; Pred. No. 2.7e-31;  
 Matches 81; Conservative 12; Mismatches 25; Indels 3; Gaps 3;

QY 1 QVQLDSSGGGLVQAGSLRLSCAASGTFSSDLMGWRQAPGKERAVARITRGCTSY 59  
 Db 1 EVQLVESGGGLVQPGGSLRLSCAASGFTVSSNNMWVQAQGLKESVYLVSSGSSYY 60

QY 60 ADSVKGRTISRDNKNTMYLQMNSLKEPEDTAVYYCCNARSNYDRSGWDYWGCGTQTVSS 119  
 Db 61 PDSVKGRTISRDNKNTMYLQMNSLKEPEDTAVYYC--ARDGIDYVGFV-YWGCGTLTVSS 118

QY 120 S 120  
 Db 119 A 119

## RESULT 4

Q9UL88 PRELIMINARY; PRT; 131 AA.

AC Q9UL88;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
 DE Myosin-reactive immunoglobulin heavy chain variable region  
 (Fragment).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98277139; PubMed=9614934;  
 RA Wu X., Liu B., Van der Werwe P.L., Kalis N.N., Berney S.M.,  
 RA Young D.C.;  
 RT "Myosin-reactive autoantibodies in rheumatic carditis and normal  
 fetus";  
 RL Clin. Immunol. Immunopathol. 87:184-192(1998).  
 DR EMBL; AF035026; AAD56262.1; -.  
 DR PIR; S21205; S21205.  
 DR HSSP; P01810; 2F84.  
 DR InterPro; IPR007110; Ig-like.  
 DR InterPro; IPR003596; Ig\_v.  
 DR Pfam; PF00047; Ig; 1.  
 DR SMART; SM00406; IGV; 1.  
 DR PROSITE; PS50835; IG\_LIKE; 1.  
 FT NON\_TER  
 FT NON\_TER  
 SQ SEQUENCE 131 AA; 14142 MW; 96E7D668E75DEAO CRC64;

Query Match 55.8%; Score 379.5; DB 4; Length 131;  
 Best Local Similarity 59.6%; Pred. No. 3e-31;  
 Matches 81; Conservative 14; Mismatches 20; Indels 21; Gaps 3;

QY 1 QVQLDSSGGGLVQAGSLRLSCAASGTFSSDLMGWRQAPGKERAVARITRGCTSY 57  
 Db 1 EVQLVESGGGLVQPGGSLRLSCAASGFTVSSNNMWVQAQGLKESVYLVSSGSSYY 60  
 QY 58 SYADSVKGRFTISRDNKNTMYLQMNSLKEPEDTAVYYC-----NARRSNYDR 104  
 Db 61 DYAAPVKGRLISRDNKNTMYLQMNSLKEPEDTAVYYCTGTGIMTIVITSSKRTSF-- 118  
 QY 105 SWGDIWGCGTQTVSS 120

Db 119 ---EYMGGLVTVSS 131

## RESULT 5

ID Q9UL90 PRELIMINARY; PRT; 113 AA.

DT 01-MAY-2000 (TrEMBLrel. 13, Created)

DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Myosin-reactive immunoglobulin heavy chain variable region (Fragment).

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Euthera; Primates; Catarrhini; Homiidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=9827139; PubMed=9614934;

RA Wu X., Liu B., Van der Werwe P.L., Kalis N.N., Berney S.M.,

RT "Myosin-reactive autoantibodies in rheumatic carditis and normal

RT fetus.";

RL Clin. Immunol. Immunopathol. 87:184-192(1998).

DR EMBL; AF035024; AAD56260.1; -.

DR HSSP; P01772; 2FB4.

DR InterPro: IPR007110; IG-like.

DR InterPro: IPR003596; IG\_v.

DR Pfam; PF00047; IG; 1.

DR SMART; SM00406; IGV; 1.

DR PROSITE; PSS0835; IG\_LIKE; 1.

FT NON\_TER 1 1

FT NON\_TER 113 113

SQ SEQUENCE 113 AA; 12437 MW; ED57FDD19086D07F CRC64;

Query Match 55.4%; Score 376.5; DB 4; Length 113;

Best Local Similarity 65.0%; Pred. No. 5.1e-31;

Matches 83; Conservative 8; Mismatches 21; Indels 9; Gaps 3;

QY 1 OVQLDSSGGGLVQAGSLRLSCAASGSIFFSSDLMGWYRQAPGKERAVARTITGGTTS-Y 59

Db 1 EVQLVESGGGVQPGSSLRISCAASGFTFSYGMHWYRQAPGKLEWVAITRDGSKYY 60

QY 60 ADSVKGRTISRDNKNTMYLQNMSLKPEDTAVVYCNARRSNYDRSGWYGGGTQVTVS 119

Db 61 ADSVKGRTISRDNKNTMYLQNMSLKPEDTAVVYCNARRSNYDRSGWYGGGTQVTVS 112

QY 120 S 120

Db 113 S 113

## RESULT 6

ID Q8WUK1 PRELIMINARY; PRT; 613 AA.

DT 01-MAR-2002 (TrEMBLrel. 20, Created)

DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Hypothetical protein.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Euthera; Primates; Catarrhini; Homiidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=tonsil;

RA Strauberg R.;

RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; BC020240; AA020240.1; -.

DR PIR; P0120; P0120.

DR PIR; S1590; S1590.

DR InterPro; IPR007110; IG-like.

DR InterPro; IPR003006; IG\_MHC.

DR InterPro; IPR003596; IG\_v.

DR Pfam; PF00047; IG; 5.

DR SMART; SM00406; IGV; 1.

DR PROSITE; PSS0835; IG\_LIKE; 5.

DR PROSITE; PSS0290; IG\_MHC; 3.

KM Hypothetical protein

SQ SEQUENCE 613 AA; 67296 MW; 60C7F5950671E315 CRC64;

Query Match 55.1%; Score 375; DB 4; Length 613;

Best Local Similarity 63.4%; Pred. No. 6.5e-30;

Matches 83; Conservative 8; Mismatches 26; Indels 14; Gaps 3;

QY 1 OVQLDSSGGGLVQAGSLRLSCAASGSIFFSSDLMGWYRQAPGKERAVARTITGGTTS-Y 59

Db 20 EVQLVESGGGVQPGSSLRISCAASGFTFSYGMHWYRQAPGKLEWVAISYDGSNRY 79

QY 60 ADSVKGRTISRDNKNTMYLQNMSLKPEDTAVVYCNARRSNYDRSGWYGGGTQVTVS 113

Db 80 ADSVKGRTISRDNKNTMYLQNMSLKPEDTAVVYCNARRSNYDRSGWYGGGTQVTVS 132

QY 114 TQVTVSSAHS 124

Db 133 TWVTVSSGSAS 143

QY 114 TQVTVSSAHS 124

Db 133 TWVTVSSGSAS 143

## RESULT 7

ID Q9UL71 PRELIMINARY; PRT; 121 AA.

DT 01-MAY-2000 (TrEMBLrel. 13, Created)

DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Myosin-reactive immunoglobulin heavy chain variable region (Fragment).

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Euthera; Primates; Catarrhini; Homiidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=9827139; PubMed=9614934;

RA Wu X., Liu B., Van der Werwe P.L., Kalis N.N., Berney S.M.,

RT "Myosin-reactive autoantibodies in rheumatic carditis and normal

RT fetus.";

RL Clin. Immunol. Immunopathol. 87:184-192(1998).

DR EMBL; AF035043; AAD56279.1; -.

DR HSSP; P01772; 2FB4.

DR InterPro: IPR007110; IG-like.

DR InterPro: IPR003596; IG\_v.

DR Pfam; PF00047; IG; 1.

DR SMART; SM00406; IGV; 1.

DR PROSITE; PSS0835; IG\_LIKE; 1.

FT NON\_TER 1 1

FT NON\_TER 121 121

SQ SEQUENCE 121 AA; 13154 MW; 2F045CCFA5D50736 CRC64;

Query Match 54.9%; Score 373.5; DB 4; Length 121;

Best Local Similarity 65.0%; Pred. No. 1.1e-30;

Matches 80; Conservative 13; Mismatches 25; Indels 5; Gaps 3;

QY 1 OVQLDSSGGGLVQAGSLRLSCAASGSIFFSSDLMGWYRQAPGKERAVARTITGGTTS-Y 59

Db 1 EVQLVESGGGVQPGSSLRISCAASGFTFSYGMHWYRQAPGKLEWVAISYDGSNRY 60

QY 60 ADSVKGRTISRDNKNTMYLQNMSLKPEDTAVVYCNARR--SNYDRSGWYGGGTQVTVS 117

Db 61 ADSVKGRTISRDNKNTMYLQNMSLKPEDTAVVYCNARR--DINQGGTQVTVS 118

QY 118 VSS 120

Db 119 VSS 121

```
RESULT 8
O9509 PRELIMINARY; PRT; 147 AA.
AC O9509;
BT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DE 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE VH3 protein (Fragment).
GN VH3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96071149; PubMed=7475288;
RA Cao J., Vescio R.A., Rettig M.B., Hong C.H., Kim A., Lee J.C.,
RA Lichtenstein A.K., Berenson J.R.;
RT "A CD10-positive subset of malignant cells is identified in multiple
RL leukemia 9:1948-1953 (1995).
DR EMBL; S80860; AAD14339.1; -.
DR HSSP; P01772; 2FB4.
DR GO; GO:0005887; C:integral to plasma membrane; NAS.
DR GO; GO:0016489; F:immunoglobulin receptor activity; NAS.
DR GO; GO:0016066; F:cellular defense response (sensu Vertebrata); NAS.
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003596; IG_V.
DR Pfam; PF00047; 1g; 1.
DR SMART; SMO0406; IGV; 1.
DR PROSITE; PS50835; IG_LIKE; 1.
DR NON_TER 1
FT SEQUENCE 147 AA; 15768 MW; 8489FCMAA7BC925C CRC64;
SQ

Query Match
Best Local Similarity 54.9%; Score 373.5; DB 4; Length 147;
Matches 85; Conservative 12; Mismatches 29; Indels 9; Gaps 4;

OY 1 QYOLDSGGGLVQAGSRLSCAASGSIFFSSDLMGWYRQAPGKEREAVARITRGTSY 59
DB 1 QYHVESGGGVVQPGKSLRLSCAASGFTFTSYGMSWVQAQKGLDWALISYDGSYQY 60
OY 60 ADVYKGRFTISDNKNTMYLQMSLKEPDYAVYYCNARSRNYDRSWG-----DYWGQG 113
DB 61 AGSVKGRFTISDNKNTMYLQMSLKEPDYAVYYCNARSRNYDRSWG-----DYWGQG 119
OY 114 QYVYSSAHHSDEPS 128
DB 120 TLVTVSSA-STKGPS 133

RESULT 9
O9HCC1 PRELIMINARY; PRT; 112 AA.
AC O9HCC1;
BT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DE 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Single chain Fv (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX Kikuchi M., Takega C., Tsujimoto Y., Asada S., Nagata K.;
RT "An antibody fragment2A3 specific for native lysocyme: isolation from a
RL human synthetic phage display library and characterization."
DR EMBL; AB049315; BAB16829.1; -.
DR HSSP; P01772; 2FB4.
```

```
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003596; IG_V.
DR Pfam; PF00047; 1g; 1.
DR SMART; SMO0406; IGV; 1.
DR PROSITE; PS50835; IG_LIKE; 1.
FT NON_TER 1
FT NON_TER 1
SQ SEQUENCE 112 AA; 12243 MW; 24F1A45ECB84788 CRC64;

Query Match
Best Local Similarity 54.6%; Score 371; DB 4; Length 112;
Matches 77; Conservative 10; Mismatches 24; Indels 6; Gaps 2;

OY 1 QYOLDSGGGLVQAGSRLSCAASGSIFFSSDLMGWYRQAPGKEREAVARIT-RGTSY 59
DB 1 EVQLVESGGGVVQPGKSLRLSCAASGFTFTSYGMSWVQAQKGLDWALISYDGSYQY 60
OY 60 ADVYKGRFTISDNKNTMYLQMSLKEPDYAVYYCNARSRNYDRSWGDIYGGQTV 116
DB 61 ADVYKGRFTISDNKNTMYLQMSLKEPDYAVYYCNARSRNYDRSWGDIYGGQTV 112

RESULT 10
O9UL91 PRELIMINARY; PRT; 118 AA.
AC O9UL91;
BT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Myosin-reactive immunoglobulin heavy chain variable region
DE (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98277139; PubMed=9614934;
RA Wu X., Liu B., Van der Merwe P.L., Kalis N.N., Berney S.M.,
RA Young D.C.;
RT "Myosin-reactive autoantibodies in rheumatic carditis and normal
RT fetus."
RL Clin. Immunol. Immunopathol. 87:184-192 (1998).
DR EMBL; AF035023; AAD56259.1; -.
DR PIR; S21205; S21205.
DR HSSP; P01772; 2FB4.
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003596; IG_V.
DR Pfam; PF00047; 1g; 1.
DR SMART; SMO0406; IGV; 1.
DR PROSITE; PS50835; IG_LIKE; 1.
FT NON_TER 1
FT NON_TER 1
SQ SEQUENCE 118 AA; 12843 MW; D0633949F2AC149D CRC64;

Query Match
Best Local Similarity 54.3%; Score 369; DB 4; Length 118;
Matches 81; Conservative 10; Mismatches 25; Indels 4; Gaps 3;

OY 1 QYOLDSGGGLVQAGSRLSCAASGSIFFSSDLMGWYRQAPGKEREAVARITRG-GTTSY 59
DB 1 EVQLVESGGGVVQPGKSLRLSCAASGFTFTSYGMSWVQAQKGLDWALISYDGSYQY 60
OY 60 ADVYKGRFTISDNKNTMYLQMSLKEPDYAVYYCNARSRNYDRSWGDIYGGQTV 119
DB 61 ADVYKGRFTISDNKNTMYLQMSLKEPDYAVYYCNARSRNYDRSWGDIYGGQTV 117

RESULT 11
O8NSK4 PRELIMINARY; PRT; 499 AA.
AC O8NSK4;
BT 01-OCT-2002 (TREMBlrel. 22, Created)
```



DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)  
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
 DE Hypochemical protein.  
 OS Homo sapiens (human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 [1]\_  
 SEQUENCE FROM N.A.  
 RP TRISUB=Blood;  
 RC Strausberg R.;  
 RA Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.  
 RL EMBL; BC032249; AA032249.1; -.  
 DR InterPro; IPR003599; IG.  
 DR InterPro; IPR007110; IG-like.  
 DR InterPro; IPR003597; IG\_c1.  
 DR InterPro; IPR003006; IG\_MHC.  
 DR InterPro; IPR003596; IG\_V.  
 DR Pfam; PF00047; IG\_4.  
 DR SMART; SM00409; IG\_4.  
 DR SMART; SM00407; IGc1\_2.  
 DR SMART; SM00406; IG\_V\_1.  
 DR PROSITE; PS00835; IG\_LIKE; 4.  
 DR PROSITE; PS00290; IG\_MHC\_1.  
 KM Hypochemical protein  
 SQ SEQUENCE 499 AA; 53376 MW; 93A5C89582054F32 CRC64;

Query Match	54.0%;	Score 367.5;	DB 4;	Length 499;
Best Local Similarity	60.2%;	Pred. No. 2.9e-29;		
Matches 77;	Conservative 14;	Mismatches 30;	Indels 7;	Gaps 2

QY	1	QVQLDQSGGGLVQVAGSGLRLSCAASGSLFSSDILNGWVROAPGKERELVAIIT-RTGATIEY	59
	:		:
	:		:
	:		:
Db	20	EVQLVESGGGVVRRPGGSLRLSCATGSPFDDSGASWVROAPGKLEWVISINNGSTNY	79
	:		:
	:		:
	:		:
QY	60	ABSVKRPTISRDNKNTWYLOMNSLKEPDPAVYYCAARSSNDRS-----WGCVWGQ	113
	:		:
	:		:
	:		:
Db	80	ABSVKRFITISDNKNSLYLOMNSLRPDYALYYCARDPKYGSGSGCIYIMDWGK	139
	:		:
	:		:
	:		:
QY	114	QVTVSSA	121
	:		:
	:		:
	:		:
Db	140	TVTVSSA	147
	:		:
	:		:
	:		:

## RESULT 12

Q99KA4	PRELIMINARY;	PRT;	487	AA
ID	Q99KA4			
TC	Q99KA4			

Query Match	54.0%;	Score 367;	DB 11;	Length 487;
-------------	--------	------------	--------	-------------

Best Local Similarity 61.9%; Pred. No. 3.2e-29;  
Matches 78; Conservative 12; Mismatches 28; Indels 8; Gaps 3;

**QY** I QVQLDSDGGGGLVQAAGSGLRRLSCAASSGITFSSDLGMGYRPAQPGKERAANAVALIRRGST-**TSY** 59  
:  
:  
:  
**Dd** 20 EQVLVESGGGLVKPQGSLIKLSCASGDTFSSYSMSWNRQPPELRLEAWATLIDGGSIYY 79

Db 80 PDNTRGRFTLSRDNAKNLNYLQMSHLSKSEDTAMYYCARDMGGSFYGGYSRF--DYMGQGT 137

QY	115 QVTWSS	120
	:	
Db	138 TITWSS	143

## RESULT 13

ID	Q8WU38	PRELIMINARY;	PRT;	573 AA
----	--------	--------------	------	--------

DT	01-MAR-2002	(TREMBLrel. 20, Created)
DT	01-MAR-2002	(TREMBLrel. 20, Last sequence update)
DT	01-OCT-2003	(TREMBLrel. 25, Last annotation update)
DE	Hypothetical protein.	
OS	Homo sapiens (Human).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.	
OX	[1]	
RN	NCBI_Taxid=9606;	
RP	SEQUENCE FROM N.A.	
KC	TISSUE=tomsl;	
RA	Straussberg R.;	
RL	Submitted (JAN-2002) to the EMBL/GenBank/DDBJ databases.	
DR	EMBL BC021276; AAH31276.1; -	
DR	PIR; S21205; S21205-	
DR	InterPro; IPR007110; IG-like.	
DR	InterPro; IPR003006; IG_MHC.	
DR	InterPro; IPR003596; IG_v.	
DR	Pfam; PF00047; Ig; 4.	
DR	SMART; SMC0406; IGV; 1.	
DR	PROSITE; PS50835; IG_LIKE; 4.	
DR	PROSITE; PS00290; IG_MHC; 2.	
KM	Hypothetical protein.	
SQ	SEQUENCE 573 AA; 62967 MW; F0D0T2344033AC530 CRC64;	

Query Match	53.5%	Score 364;	DB 4;	Length 573;
-------------	-------	------------	-------	-------------

Best Local Similarity 64.0%; Pred. No. 8.1e-29;  
Matches 80; Conservative 11; Mismatches 30; Indels 4; Gaps 3;

QY	1	YVQVLQDSSGGLVQAGSGSLRSLSCAASGSIFSSDLDMGWYRQAPGKESEAVART-RCGTTSY	59
	:		:
Db	20	EVQLVDSGGGLVQGRSLRLSCAASGTEPDYAMHWRAAPGKGLFWYSGISINSGSIGY	79
	:		:
QY	60	ADSVKGRFTISRDNKNTMTLQNNSLKPEDYAVYC--NARRSNYDRSMG--DYWGQGTQV	116
	:		:
QY	80	ADSVKGRFTISRDNKNSLYLQNNSLRAEDTALYYCAKKGSGSYIGYYGMDVMWGQGTTV	139
	:		:
Db	140	TVSSA 144	
	:		:
QY	117	TVSSA 121	
	:		:
Db	140	TVSSA 144	
	:		:
RESULT 14			
Q9UL93			
ID	Q9UL93	PRELIMINARY;	PRT; 116 AA.
AC	Q9UL93;		
DT	01-MAY-2000 (TREMBLrel. 13, Created)		
DT	01-MAY-2000 (TREMBLrel. 13, Last sequence update)		
DT	01-OCT-2003 (TREMBLrel. 25, Last annotation update)		
DE	Myosin-reactive immunoglobulin heavy chain variable region (Fragment).		
DE	Homio sapiens (Human).		
OS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
CC			

OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=98277139; PubMed=9614934;  
RA Wu X., Liu B., Van der Merwe P.L., Kallie N.N., Berney S.M.,  
RT "Myosin-reactive autoantibodies in rheumatic carditis and normal  
fetus";  
RL Clin. Immunol. Immunopathol. 87:184-192(1998).  
DR EMBL; AF035021; AAD56257.1; -.  
DR PIR; P0120; P0120.  
DR HSSP; P01772; 2F84.  
DR InterPro; IPR007110; Ig-like.  
DR InterPro; IPR003596; Ig\_V.  
DR Pfam; PF00047; Ig\_1.  
DR SMART; SM00406; IGV; 1.  
DR PROSITE; PS00835; IG\_LIKE; 1.  
FT NON\_TER 1  
FT TER 1  
SQ SEQUENCE 116 AA; 12434 MW; 0DA0348154DD6061 CRC64;  
  
Query Match 53.2%; Score 361.5; DB 4; Length 116;  
Best Local Similarity 64.0%; Pred. No. 1.8e-29;  
Matches 80; Conservative 9; Mismatches 21; Indels 15; Gaps 3;  
  
QY 2 VOLODSGGGLVQAGSLRLSCAASGIFSSDLMGMWYRQAPGKEREAARITRGSTT-SY 60  
DB 1 VOLVSSGGGLVQAGSLRLSCAASGIFSSDLMGMWYRQAPGKEREAARITRGSTT-SY 60  
QY 61 DSVKGRFTISRDNAKNTLYQMNSLKPEDTAVYYCNARNSNYDRSWG----DYWGQGTQ 115  
DB 61 DSVKGRFTISRDNSNKTLYQMNSLKPEDTAVYYC-----AGGGGLGLGYWGQGT 111  
QY 116 VTVSS 120  
DB 112 VTVSS 116  
  
RESULT 15  
Q91Z05 PRELIMINARY; PRT; 473 AA.  
AC Q91Z05;  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Hypothetical protein.  
GN AU044919.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Strausberg R.;  
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BC010327; AAH10327.1; -.  
DR WGD; WGI:2144967; AU044919.  
DR GO; GO:0005489; Folate transporter activity; IEA.  
DR GO; GO:0006118; Pilection transport; IEA.  
DR InterPro; IPR00345; Cytochrome\_BS.  
DR InterPro; IPR007110; Ig-like.  
DR InterPro; IPR003006; Ig\_MHC.  
DR InterPro; IPR003596; Ig\_V.  
DR Pfam; PF00047; Ig\_3.  
DR SMART; SM00406; IGV; 1.  
DR PROSITE; PS00190; CYTOCHROME\_C; 1.  
DR PROSITE; PS00835; IG\_LIKE; 4.  
DR PROSITE; PS00290; IG\_MHC; 1.  
KM Hypothetical protein  
SQ SEQUENCE 473 AA; 51946 MW; CF625F008932AF12 CRC64;

Query Match

52.9%; Score 360; DB 11; Length 473;

Best Local Similarity 65.6%; Pred. No. 1.6e-28;  
Matches 80; Conservative 10; Mismatches 28; Indels 4; Gaps 3;  
  
QY 1 VOLODSGGGLVQAGSLRLSCAASGIFSSDLMGMWYRQAPGKEREAARITRGSTT-SY 59  
DB 20 VOLVSSGGGLVQAGSLRLSCAASGIFSSDLMGMWYRQAPGKEREAARITRGSTT-SY 79  
QY 60 ADSVKGRTISRDNAKNTLYQMNSLKPEDTAVYYCNARNSNYDRSWGDYWGQGTQVTVS 119  
DB 80 ADSVKGRTISRDNAKNTLYQMNSLKPEDTAVYYC--ARELMLRRI--DYWGQGTITVS 136  
QY 120 SA 121  
DB 137 SA 138  
  
Search completed: October 5, 2004, 08:13:43  
Job time: 69.2824 secs



XX The patent discloses antibodies or their fragments comprising a heavy  
CC chain variable domain (VH) derived from an immunoglobulin naturally  
CC devoid of light chains specific for inhibiting human dietary enzymes. The  
CC antibodies of the invention are useful for the preparation of medicaments  
CC or food for inhibiting the activity of one or more human dietary enzymes  
CC especially human pancreatic lipase (HPL) or human gastric lipase (HGL)  
CC which are useful for the cosmetic control of body weight of human beings.  
CC The present peptide sequence is HPL inhibiting VH fragment, HPL #11 from  
CC llama (camelid) species  
XX

Sequence 129 AA:

Query Match 100.0%; Score 680; DB 4; Length 129;  
Best Local Similarity 100.0%; Pred. No. 6.7e-54;  
Matches 129; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QVQLDSGGGLVQAGSLRLSCAASGIFSSDLMGMYRQAPGKERAVARIRGGTSTYA 60  
DB 1 QVQLDSGGGLVQAGSLRLSCAASGIFSSDLMGMYRQAPGKERAVARIRGGTSTYA 60  
QY 61 DSVKGRFTISRDNKNTMYLQWNSLKPEDTAVYYCNARRSNYDRSGWDYWGCGTQVTVSS 120  
DB 61 DSVKGRFTISRDNKNTMYLQWNSLKPEDTAVYYCNARRSNYDRSGWDYWGCGTQVTVSS 120  
QY 121 AHSEDPSS 129  
DB 121 AHSEDPSS 129

RESULT 2  
AAE10558  
ID AAE10558 standard; peptide; 129 AA.

XX AAE10558;  
AC  
XX  
XX 10-DEC-2001 (first entry)  
DT  
XX  
XX HPL inhibiting VH fragment, HPL #22 from llama species.  
DE

XX llama antibody; camelid; anorectic; heavy chain variable domain; VH;  
KW human dietary enzyme inhibitor; medicament; human pancreatic lipase; HPL;  
KW food; human gastric lipase; HGL; cosmetic control; body weight.

OS Lama sp.

XX  
XX Key Location/Qualifiers  
FT 31..35  
FT Region /label= CDR1  
FT /note= "Complementarity determining region 1"

FT Region /label= CDR2  
FT /note= "Complementarity determining region 2"

FT Region /label= CDR3  
FT /note= "Complementarity determining region 3"

EP1134231-A1.

XX 19-SEP-2001.

XX 20-FEB-2001; 2001EP-00200703.

XX 14-MAR-2000; 2000EP-00200930.

XX (UNIL ) UNILEVER NV.  
PA (UNIL ) UNILEVER PLC.

XX Bezemer S, Van De Burg M, De Haard JMW, Tareilus E;

XX WPI; 2001-572718/65.

XX New antibody or its fragments for inhibiting human dietary enzymes,  
PT

PT useful for cosmetic control of body weight of human beings, comprises  
PT heavy chain variable domain derived from immunoglobulin naturally devoid  
PT of light chains.  
XX  
XX Example 2, Page 10; 37p; English.

XX The patent discloses antibodies or their fragments comprising a heavy  
CC chain variable domain (VH) derived from an immunoglobulin naturally  
CC devoid of light chains specific for inhibiting human dietary enzymes. The  
CC antibodies of the invention are useful for the preparation of medicaments  
CC or food for inhibiting the activity of one or more human dietary enzymes  
CC especially human pancreatic lipase (HPL) or human gastric lipase (HGL)  
CC which are useful for the cosmetic control of body weight of human beings.  
CC The present peptide sequence is HPL inhibiting VH fragment, HPL #22 from  
CC llama (camelid) species  
XX

Sequence 129 AA:

Query Match 73.7%; Score 501; DB 4; Length 129;  
Best Local Similarity 77.2%; Pred. No. 1.1e-37;  
Matches 98; Conservative 9; Mismatches 20; Indels 0; Gaps 0;

QY 1 QVQLDSGGGLVQAGSLRLSCAASGIFSSDLMGMYRQAPGKERAVARIRGGTSTYA 60  
DB 1 QVQLDSGGGLVQAGSLRLSCAASGIFSSDLMGMYRQAPGKERAVARIRGGTSTYA 60  
QY 61 DSVKGRFTISRDNKNTMYLQWNSLKPEDTAVYYCNARRSNYDRSGWDYWGCGTQVTVSS 120  
DB 61 DSVKGRFTISRDNKNTMYLQWNSLKPEDTAVYYCNARRSNYDRSGWDYWGCGTQVTVSS 120  
QY 121 AHSEDP 127  
DB 121 EPTKPKP 127

RESULT 3  
AAE10553  
ID AAE10553 standard; peptide; 129 AA.

XX AAE10553;  
AC  
XX  
XX 10-DEC-2001 (first entry)  
DT  
XX  
XX HPL inhibiting VH fragment, HPL #13 from llama species.  
DE

XX llama antibody; camelid; anorectic; heavy chain variable domain; VH;  
KW human dietary enzyme inhibitor; medicament; human pancreatic lipase; HPL;  
KW food; human gastric lipase; HGL; cosmetic control; body weight.

OS Lama sp.

XX  
XX Key Location/Qualifiers  
FT 31..35  
FT Region /label= CDR1  
FT /note= "Complementarity determining region 1"

FT Region /label= CDR2  
FT /note= "Complementarity determining region 2"

FT Region /label= CDR3  
FT /note= "Complementarity determining region 3"

EP1134231-A1.

XX 19-SEP-2001.

XX 20-FEB-2001; 2001EP-00200703.

XX 14-MAR-2000; 2000EP-00200930.

XX (UNIL ) UNILEVER NV.  
PA (UNIL ) UNILEVER PLC.

PI Bezemer S, Van De Burg M, De Haard JWM, Tareilus E;  
 XX WPI; 2001-572718/65.  
 DR  
 XX  
 XX New antibody or its fragments for inhibiting human dietary enzymes,  
 PT useful for cosmetic control of body weight of human beings, comprises  
 PT heavy chain variable domain derived from immunoglobulin naturally devoid  
 PT of light chains.  
 XX  
 XX Example 2; Page 9; 37pp; English.  
 PS  
 XX The patent discloses antibodies or their fragments comprising a heavy  
 CC chain variable domain (VHH) derived from an immunoglobulin naturally  
 CC devoid of light chains specific for inhibiting human dietary enzymes. The  
 CC antibodies of the invention are useful for the preparation of medicaments  
 CC or food for inhibiting the activity of one or more human dietary enzymes  
 CC especially human pancreatic lipase (HPL) or human gastric lipase (HGL)  
 CC which are useful for the cosmetic control of body weight of human beings.  
 CC The present peptide sequence is HPL inhibiting VHH fragment, HPL #13 from  
 CC llama (camelid) species  
 XX  
 XX Sequence 129 AA;  
 SQ  
 Query Match 73.5%; Score 500; DB 4; Length 129;  
 Best Local Similarity 75.6%; Pred. No. 1.3e-37;  
 Matches 96; Conservative 11; Mismatches 20; Indels 0; Gaps 0;  
 QY 1 QVQLDSGGGLVQAGSLRLSCAASGIFSSDLMGWYRQAQKREAVARITRGTTSYA 60  
 Db 1 QVQLDSGGGLVQAGSLRLSCAASGIFSSDLMGWYRQAQKREAVARITRGTTSYA 60  
 QY 61 DSVKGRFTISRDNKNTMYLQNNSLKPEDTAVYYCNARSRYSMDYWGQGTQVTVSS 120  
 Db 61 DSAKGRFTISKDNKNTMYLQNNSLKPEDTAVYYCNALITRWKDSVNDYWGQGTQVTVSS 120  
 QY 121 AHSSEDP 127  
 Db 121 EPKTPKP 127  
 DB  
 RESULT 4  
 AAE10560  
 ID AAE10560 standard; peptide; 130 AA.  
 AC AAE10560;  
 XX  
 XX 10-DEC-2001 (first entry)  
 DT  
 XX HGL inhibiting VHH fragment, HGL #1 from llama species.  
 DE  
 XX Llama antibody; camelid; anorectic; heavy chain variable domain; VHH;  
 KM human dietary enzyme inhibitor; medicament; human pancreatic lipase; HPL;  
 KW food; human gastric lipase; HGL; cosmetic control; body weight.  
 XX  
 XX Llama sp.  
 OS  
 XX  
 XX Key Location/Qualifiers  
 FH 31..35  
 FT Region /label= CDR1  
 FT /note= "Complementarity determining region 1"  
 FT 50..64  
 FT Region /label= CDR2  
 FT /note= "Complementarity determining region 2"  
 FT 98..110  
 FT Region /label= CDR3  
 FT /note= "Complementarity determining region 3"  
 XX  
 XX EPI134231-A1.  
 PD 19-SEP-2001.  
 XX  
 XX 20-FEB-2001; 2001EP-00200703.  
 PF  
 XX

PR 14-MAR-2000; 2000EP-00200930.  
 XX  
 XX (UNIT ) UNILEVER NV.  
 PA (UNIT ) UNILEVER PLC.  
 XX  
 XX Bezemer S, Van De Burg M, De Haard JWM, Tareilus E;  
 PI WPI; 2001-572718/65.  
 DR  
 XX  
 XX New antibody or its fragments for inhibiting human dietary enzymes,  
 PT useful for cosmetic control of body weight of human beings, comprises  
 PT heavy chain variable domain derived from immunoglobulin naturally devoid  
 PT of light chains.  
 XX  
 XX Example 4; Page 13; 37pp; English.  
 PS  
 XX The patent discloses antibodies or their fragments comprising a heavy  
 CC chain variable domain (VHH) derived from an immunoglobulin naturally  
 CC devoid of light chains specific for inhibiting human dietary enzymes. The  
 CC antibodies of the invention are useful for the preparation of medicaments  
 CC or food for inhibiting the activity of one or more human dietary enzymes  
 CC especially human pancreatic lipase (HPL) or human gastric lipase (HGL)  
 CC which are useful for the cosmetic control of body weight of human beings.  
 CC The present peptide sequence is HGL inhibiting VHH fragment, HGL #1 from  
 CC llama (camelid) species  
 XX  
 XX Sequence 130 AA;  
 SQ  
 Query Match 73.3%; Score 498.5; DB 4; Length 130;  
 Best Local Similarity 75.6%; Pred. No. 1.8e-37;  
 Matches 102; Conservative 6; Mismatches 16; Indels 11; Gaps 2;  
 QY 1 QVQLDSGGGLVQAGSLRLSCAASGIFSSDLMGWYRQAQKREAVARITRGTTSYA 60  
 Db 1 QVQLDSGGGLVQAGSLRLSCAASGIFSSDLMGWYRQAQKREAVARITRGTTSYA 60  
 QY 61 DSVKGRFTISRDNKNTMYLQNNSLKPEDTAVYYCNARSRYSMDYWGQGTQVTVSS 114  
 Db 61 DSVKGRFTISRDNKNTMYLQNNSLKPEDTAVYYCNARSRYSMDYWGQGTQVTVSS 115  
 QY 115 QVTVSSAHSSEDPSS 129  
 Db 116 QVTVSSAHSSEDPSS 130  
 DB  
 RESULT 5  
 AAE10562  
 ID AAE10562 standard; peptide; 124 AA.  
 AC AAE10562;  
 XX  
 XX 10-DEC-2001 (first entry)  
 DT  
 XX HGL inhibiting VHH fragment, HGL #8 from llama species.  
 DE  
 XX Llama antibody; camelid; anorectic; heavy chain variable domain; VHH;  
 KM human dietary enzyme inhibitor; medicament; human pancreatic lipase; HPL;  
 KW food; human gastric lipase; HGL; cosmetic control; body weight.  
 XX  
 XX Llama sp.  
 OS  
 XX  
 XX Key Location/Qualifiers  
 FH 31..35  
 FT Region /label= CDR1  
 FT /note= "Complementarity determining region 1"  
 FT 50..64  
 FT Region /label= CDR2  
 FT /note= "Complementarity determining region 2"  
 FT 98..104  
 FT Region /label= CDR3  
 FT /note= "Complementarity determining region 3"  
 XX  
 XX EPI134231-A1.  
 PD  
 XX  
 XX

```

XX 19-SEP-2001.
PD 20-FEB-2001; 2001EP-00200703.
XX 14-MAR-2000; 2000EP-00200930.
XX (UNIL ) UNILEVER NV.
PA (UNIL ) UNILEVER PLC.
XX Bezemer S, Van De Burg M, De Haard JWM, Tareilus E;
PI WPI; 2001-572718/65.
XX
XX New antibody or its fragments for inhibiting human dietary enzymes,
PT useful for cosmetic control of body weight of human beings, comprises
PT heavy chain variable domain derived from immunoglobulin naturally devoid
PT of light chains.
XX
XX Example 4; Page 13; 37pp; English.
XX
XX The patent discloses antibodies or their fragments comprising a heavy
CC chain variable domain (VH) derived from an immunoglobulin naturally
CC devoid of light chains specific for inhibiting human dietary enzymes. The
CC antibodies of the invention are useful for the preparation of medicaments
CC or food for inhibiting the activity of one or more human dietary enzymes
CC especially human pancreatic lipase (HPL) or human gastric lipase (HGL)
CC which are useful for the cosmetic control of body weight of human beings.
CC The present peptide sequence is HGL inhibiting VH fragment, HGL #8 from
CC llama (camelid) species
XX
XX Sequence 124 AA:
SQ
Query Match 72.6%; Score 493.5; DB 4; Length 124;
Best Local Similarity 75.2%; Pred. No. 4.9e-37;
Matches 97; Conservative 8; Mismatches 19; Indels 5; Gaps 1;
QY 1 QVQLDSSGGLVQAGSLRLSCAASGIFSSDIIMGWYRQAPGKERAVARITRGTTSYA 60
DB 1 QVQLDSSGGLVQAGSLRLSCAASGIFSSDIIMGWYRQAPGKERAVARITRGTTSYA 60
QY 61 DSYKGRFTISRDNKNTMYLQNNSLKPEDTAVYYCNARSNYDRSGWDYWGCGTQVTS 120
DB 61 DSYKGRFTISRDNKNTMYLQNNSLKPEDTAVYYCNARSNYDRSGWDYWGCGTQVTS 115
QY 121 AHSSEDPSS 129
DB 116 AHSSEDPSS 124
RESULT 6
AAE05284
ID AAE05284 standard; protein; 150 AA.
XX
AC AAE05284;
XX
DT 18-SEP-2001 (first entry)
XX
DE Anti-GUS (Clone18) VH region attached with myc and his6 tag.
XX
KW Heavy chain immunoglobulin; pathogen resistance; metabolism modulator;
KW passive immunisation; heavy chain variable domain; VH; anti-GUS antibody;
KW beta-glucuronidase.
XX
OS Unidentified.
XX
XX EPI118669-A2.
XX
XX 25-JUL-2001.
XX
XX 08-DEC-2000; 2000EP-00310997.
XX
XX 17-DEC-1999; 99EP-00310188.

```

```

XX (UNIL ) UNILEVER PLC.
PA (UNIL ) UNILEVER NV.
XX Frenken LGJ, Van Der Logt CPE, Jobling SA, Teh Y;
PI WPI; 2001-427157/46.
XX N-PSDB; AAD10055.
XX
XX Modifying a plant to produce an antibody useful for increasing pathogen
PT resistance or to modulate metabolism comprises introducing a DNA sequence
PT encoding a heavy chain immunoglobulin linked to a peptide that targets a
PT cellular compartment.
XX
XX Example 1; Fig 14; 81pp; English.
XX
XX The present invention relates to a method for modifying a plant to
CC produce an antibody or an active fragment or derivative, or a protein
CC functional equivalent, in a cellular compartment comprising introducing a
CC DNA sequence encoding a heavy chain immunoglobulin, where the DNA is
CC linked to promoters and provided with an additional sequence encoding a
CC peptide capable of targeting heavy chain immunoglobulin to a cellular
CC compartment. The method is used for producing a heavy chain
CC immunoglobulin or an active fragment or derivative, or a protein that is
CC functionally equivalent for increasing the pathogen resistance in a plant
CC or to modulate metabolism in a plant. Under some circumstances it may be
CC desirable to retain the antibody product with the plant rather than
CC extracting and isolating the product. In particular, edible selected
CC antigens may be used in a method of passively immunising an animal, the
CC preferably human, against the antigen, e.g., pathogenic organisms. The
CC present sequence is an anti-GUS antibody (denoted Clone18) heavy chain
CC variable domain (VH) attached to peptide linkers, myc and his6 tag. The
CC GUS is beta-glucuronidase
XX
XX Sequence 150 AA:
SQ
Query Match 72.1%; Score 490; DB 4; Length 150;
Best Local Similarity 75.4%; Pred. No. 1.3e-36;
Matches 98; Conservative 8; Mismatches 14; Indels 10; Gaps 2;
QY 1 QVQLDSSGGLVQAGSLRLSCAASGIFSSDIIMGWYRQAPGKERAVARITR-GGTTSY 59
DB 4 QVQLDSSGGLVQAGSLRLSCAASGIFSSDIIMGWYRQAPGKERAVARITR-GGTTSY 63
QY 60 ADSYKGRFTISRDNKNTMYLQNNSLKPEDTAVYYCNARSNYDRSGWDYWGCGTQVTS 119
DB 64 ADSYKGRFAVSRYDAENTVYLQNNSLKPEDTAVYYCNTRA-----YWGCGTQVTS 114
QY 120 SAHSEDPSS 129
DB 115 SAHSEDPSS 124
RESULT 7
AAE05285
ID AAE05285 standard; protein; 153 AA.
XX
AC AAE05285;
XX
DT 18-SEP-2001 (first entry)
XX
DE Anti-GUS VH region attached with ER retention signal, myc and his6.
XX
KW Heavy chain immunoglobulin; pathogen resistance; metabolism modulator;
KW passive immunisation; heavy chain variable domain; VH; anti-GUS antibody;
KW beta-glucuronidase.
XX
OS Unidentified.
XX
XX EPI118669-A2.
XX
XX 25-JUL-2001.
XX
XX

```

```

PF 08-DEC-2000; 2000EP-00310997.
XX
XX 17-DEC-1999; 99EP-00310188.
XX (UNIL ) UNILEVER PLC.
XX (UNIL ) UNILEVER NV.
XX
XX Frenken LGJ, Van Der Logt CPE, Jobling SA, Teh Y;
XX WPI; 2001-427157/46.
XX N-PSDB; AAD10056.
XX
XX Modifying a plant to produce an antibody useful for increasing pathogen
XX resistance or to modulate metabolism comprises introducing a DNA sequence
XX encoding a heavy chain immunoglobulin linked to a peptide that targets a
XX cellular compartment.
XX
XX Example 1; Fig 15; 81pp; English.
XX
XX The present invention relates to a method for modifying a plant to
XX produce an antibody or an active fragment or derivative, or a protein
XX functional equivalent, in a cellular compartment comprising introducing a
XX DNA sequence encoding a heavy chain immunoglobulin, where the DNA is
XX linked to promoters and provided with an additional sequence encoding a
XX peptide capable of targeting heavy chain immunoglobulin to a cellular
XX compartment. The method is used for producing a heavy chain
XX immunoglobulin or an active fragment or derivative, or a protein that is
XX functionally equivalent for increasing the pathogen resistance in a plant
XX or to modulate metabolism in a plant. Under some circumstances it may be
XX desirable to retain the antibody product with the plant rather than
XX extracting and isolating the product. In particular, edible selected
XX antigens may be used in a method of passively immunising an animal.
XX Preferably human, against the antigen, e.g., pathogenic organisms. The
XX present sequence is an anti-GUS antibody (denoted Clone18) heavy chain
XX variable domain (VH) attached with endoplasmic reticulum (ER) retention
XX signal (SEKDEL) and peptide linkers, myc and his6 tag. The GUS is beta-
XX glucuronidase
XX
XX Sequence 153 AA;
XX
XX Query Match 72.1%; Score 490; DB 4; Length 153;
XX Best Local Similarity 75.4%; Pred. No. 1.3e-36;
XX Matches 98; Conservative 8; Mismatches 14; Indels 10; Gaps 2;
XX
XX 1 QVQLDSGGGLVQAGSRLSCAASGIFSSDLMGMVYRQAPGKERAVARITR-GGITSY 59
XX :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX 4 QVQLQSGGGGLVQAGGPIRLSCAASGRFTFSNVAVMFQAPGKEREFVAISRGGRTYY 63
XX
XX 60 ADSVYGRFTISRDNKNTMYLQMSLKPEDTAVYYCNARRSNYDRSGDYGGGTQYTVSS 119
XX :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX 64 ADSVYGRFRAVSHDYENVTYLLQMSLKPEDTAVYYCNTR-----YWGCGTQYTVSS 114
XX
XX 120 SAHSEDPSS 129
XX :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX 115 SAHSEDPSS 124
XX
XX RESULT 8
XX AAE05282
XX ID AAE05282 standard; protein; 153 AA.
XX AC AAE05282;
XX XX
XX DT 18-SEP-2001 (first entry)
XX
XX Anti-potato SBEII antibody VH region attached with myc and his6 tag.
XX
XX Potato; heavy chain immunoglobulin; pathogen resistance;
XX metabolism modulator; passive immunisation; heavy chain variable domain;
XX VH; anti-potato SBEII; starch branching enzyme; SBE A.
XX
XX Solanum tuberosum.
XX
XX Unidentified.

```

```

OS Chimeric.
XX
XX EPI118669-A2.
XX
XX 25-JUL-2001.
XX
XX 08-DEC-2000; 2000EP-00310997.
XX
XX 17-DEC-1999; 99EP-00310188.
XX
XX (UNIL ) UNILEVER PLC.
XX (UNIL ) UNILEVER NV.
XX
XX Frenken LGJ, Van Der Logt CPE, Jobling SA, Teh Y;
XX WPI; 2001-427157/46.
XX N-PSDB; AAD10053.
XX
XX Modifying a plant to produce an antibody useful for increasing pathogen
XX resistance or to modulate metabolism comprises introducing a DNA sequence
XX encoding a heavy chain immunoglobulin linked to a peptide that targets a
XX cellular compartment.
XX
XX Example 1; Fig 6; 81pp; English.
XX
XX The present invention relates to a method for modifying a plant to
XX produce an antibody or an active fragment or derivative, or a protein
XX functional equivalent, in a cellular compartment comprising introducing a
XX DNA sequence encoding a heavy chain immunoglobulin, where the DNA is
XX linked to promoters and provided with an additional sequence encoding a
XX peptide capable of targeting heavy chain immunoglobulin to a cellular
XX compartment. The method is used for producing a heavy chain
XX immunoglobulin or an active fragment or derivative, or a protein that is
XX functionally equivalent for increasing the pathogen resistance in a plant
XX or to modulate metabolism in a plant. Under some circumstances it may be
XX desirable to retain the antibody product with the plant rather than
XX extracting and isolating the product. In particular, edible selected
XX antigens may be used in a method of passively immunising an animal.
XX Preferably human, against the antigen, e.g., pathogenic organisms. The
XX present sequence is an anti-potato SBEII antibody heavy chain variable
XX domain (VH) attached to peptide linkers, myc and his6 tag. The potato
XX SBEII is a starch branching enzyme also designated SBE A
XX
XX Sequence 153 AA;
XX
XX Query Match 72.1%; Score 490; DB 4; Length 153;
XX Best Local Similarity 75.2%; Pred. No. 1.3e-36;
XX Matches 97; Conservative 9; Mismatches 21; Indels 2; Gaps 1;
XX
XX 1 QVQLDSGGGLVQAGSRLSCAASGIFSSDLMGMVYRQAPGKERAVARITRGTTSYA 60
XX :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX 1 QVQLQSGGGGLVQAGSLRLSCAASGSSNNMGMVRRAPGQRELVAITISGGSTHYV 60
XX
XX 61 DSVYGRFTISRDNKNTMYLQMSLKPEDTAVYYCNARRSNYDRSGDYGGGTQYTVSS 120
XX :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX 61 DSVYGRFTISRDNKNTMYLQMSLKPEDTAVYYCAAKRGLOAM-QVWGCGTQYTVSS 118
XX
XX 121 AHSSEDPSS 129
XX :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX 119 AHSSEDPSS 127
XX
XX RESULT 9
XX ABR82065
XX ID ABR82065 standard; protein; 507 AA.
XX AC ABR82065;
XX XX
XX DT 23-SEP-2003 (first entry)
XX
XX Immunocjugate cAb-CEA5-beta-lactamase amino acid SEQ ID NO:14.
XX
XX Camel; antibody; carcinoembryonic antigen; CEA; anti-CEA antibody;
XX

```

KM beta-lactamase; immunocjugate; cytostatic; tumour growth inhibitor;  
KM tumour; CAB-CEAS-beta-lactamase; VH region.  
XX  
OS Camelus sp.  
OS Lama glama.  
OS Synthetic.  
XX  
FT Key Location/Qualifiers  
FT Region 1.123  
FT Region /label=CAB-CEAS\_VH\_region  
FT Region 124.138  
FT Region /label=lama\_gamma2c\_hinge\_region  
FT Region 140.1501  
FT Peptide /label=Beta-lactamase  
FT Peptide 502.1507  
FT Peptide /label=6xHis\_tag  
XX  
PN WO2003055527-A2.  
XX  
PD 10-JUL-2003.  
XX  
PF 23-DEC-2002; 2002WO-EF014842.  
XX  
PR 03-JAN-2002; 2002EP-00075048.  
PR 09-JUL-2002; 2002EP-00077734.  
XX  
XX (VLA-) VLAMS INTERUNIVERSITAIR INST BIOTECHNOC.  
XX  
PI Revets H, Cortez-Retamozo V, Muyldermans S, De Baetselier P;  
XX  
XX WPI; 2003-569408/53.  
DR N-PSDB; ACP06073.  
XX  
XX  
PT New immunocjugate useful for the manufacture of a medicament to treat  
PT tumors.  
XX  
XX Claim 7; Page 31-33; 34pp; English.  
XX  
CC The present invention describes an immunocjugate devoid of a light  
CC chain. The immunocjugate binds to a tumour antigen that contains at  
CC least one variable domain of heavy chain antibody having an anti-tumour  
CC agent attached. Also described is the use of the immunocjugate for the  
CC manufacture of a medicament to treat tumours expressing a tumour marker  
CC that is recognised by the immunocjugate. The immunocjugate has  
CC cytostatic activity and can be used as a tumour growth inhibitor. It can  
CC be used for the manufacture of a medicament to treat tumours expressing a  
CC tumour marker that is recognised by the immunocjugate. The present  
CC sequence represents a camel anti-carcinogenic antigen (CEA) antibody  
CC VH region designated CAB-CEAS and beta-lactamase immunocjugate, from  
CC the present invention  
XX  
SQ Sequence 507 AA;  
Query Match 71.5%; Score 486.5; DB 7; Length 507;  
Best Local Similarity 74.2%; Pred. No. 1,1e-35;  
Matches 98; Conservative 7; Mismatches 24; Indels 3; Gaps 2;  
QY 1 QVQLDSGGGLVQAGSLRLSCAASGSIIPSDIMGWRQAPKEREAVARIT-RGGTTSY 59  
DB 1 QVQLVSGGGSVQAGSLRLSCAASGDTYGYWMGWRQAPKEREAVAINRGGGYTY 60  
QY 60 ADVKRRFTISRDNKQTYTLQNSLKPEDTAVYICNA--RSNYDRSGDYWGQGTQVT 117  
DB 61 ADVKGRFTISRDTAVTYLQNSLRLPDDTADYVCASGVIGLHEDWENYWGQGTQVT 120  
QY 118 VSSAHSSEDPSS 129  
DB 121 VSSAHSSEDPSS 132  
RESULT 10  
ABG30619  
ID ABG30619 standard; protein; 138 AA.

XX  
AC ABG30619;  
XX  
XX 21-OCT-2002 (first entry)  
XX  
XX Immunoglobulin G specific heavy chain variable domain antibody #2.  
XX  
XX Heavy chain variable domain; antibody; protein array; aging; VHH;  
XX Immunoglobulin; sticky phage-antibody; micro-panning; immune library;  
XX proteomic; mouse.  
XX  
XX Mus sp.  
XX  
XX WO200248193-A2.  
XX  
XX 20-JUN-2002.  
XX  
XX 03-DEC-2001; 2001WO-EP014471.  
XX  
XX 13-DEC-2000; 2000EP-00311142.  
XX  
XX (UNIL ) UNILEVER NV.  
XX (UNIL ) UNILEVER PLC.  
XX (UNIL ) HINDUSTAN LEVER LTD.  
XX  
XX De Haard JWM, Hermans P, Landa I, Verrips CT;  
XX  
XX WPI; 2002-583487/62.  
XX  
XX Novel protein array useful for detecting the presence of individual  
XX proteins in sample, comprises heavy-chain variable domain antibodies or  
XX antibody fragments obtainable from Camelidae.  
XX  
XX Disclosure; Fig 3; 80pp; English.  
XX  
XX The invention describes a protein array (I) comprising a number of heavy-  
XX chain variable domain antibodies or antibody fragments, obtainable from  
XX Camelidae. The method is useful for removing abundant proteins from an  
XX extract or sample which do not provide useful information on the  
XX condition of a cell or tissue in the extract or sample to be  
XX investigated. (I) is useful for detecting the presence of individual  
XX proteins in a sample, comparing the distribution of proteins in different  
XX cell types, and identification of proteins that may be of importance in  
XX determining the altered properties of cells in disease, aging or other  
XX conditions. Using a heavy-chain variable domain derived from an  
XX immunoglobulin that is naturally devoid of light chains (VHH) in (I)  
XX provides a number of advantages, such as an improvement of  
XX sensitivity/resolution in the order of 10-100 times, and detection of  
XX post-translationally modified proteins. The invention also describes a  
XX method (II) that enables the simultaneous processing of large numbers of  
XX target antigens in a controlled way. The incorporated phage-ELISA  
XX generates on-line information about the success or failure of a certain  
XX panning condition. This feature combined with the microtiter plate format  
XX allows the complete automation of the technology, based on computer-made  
XX decisions on the values of the phage-ELISA for continuation of a limited  
XX number of selections. In (II), due to the fact that many different  
XX conditions can be tested, varying amounts of input-phages can be used  
XX simultaneously in order to decrease the enrichment of sticky phage-  
XX antibodies. Micro-panning is an effective tool for selecting both native,  
XX synthetic and immune libraries on large numbers of different target  
XX molecules, enabling the generation of large panels of antibodies in short  
XX time frames needed for the generation of arrays (proteomics). This  
XX sequence represents an immunoglobulin (IgG) heavy chain variable domain  
XX antibody (VHH)  
XX  
SQ Sequence 138 AA;  
Query Match 71.2%; Score 484.5; DB 5; Length 138;  
Best Local Similarity 72.9%; Pred. No. 3.6e-36;  
Matches 102; Conservative 8; Mismatches 17; Indels 13; Gaps 4;  
QY 1 QVQLDSGGGLVQAGSLRLSCAASGSIIPSDIMGWRQAPKEREAVARIT-RGGTTSY 59  
|||||



```

Db      1 QVQLDQSGGILVEAGSGLRLSCAVSGRTSTYAMGMFQAGKEREFVAALISGGSINH 60
QY      60 ADSVKRRFTTISRPNAKNTMYTQMSLSKPEDTAVYCN-----ARKSNDRSGMD---Y 109
Db      61 ADSVKRRFTISRNNASTVYLTQMSLSKPEDTAVYCAQNPAMVGRNNAIYR--GDEYNY 118
QY      110 WGGGTQVTVSSAHSSEDPSS 129
Db      119 WGGCAQVTVSSAHSSEDPSS 138

RESULT 11
ID      AAE05288
AAE05288 standard; protein; 149 AA.
AC      AAE05288;
AD      18-SEP-2001 (first entry)
DE      Anti-potato SBEII (Clone68) VH region attached with myc and his6 tag.
DX      Potato; heavy chain immunoglobulin; pathogen resistance;
KM      metabolism modulator; passive immunisation; heavy chain variable domain;
KN      VH; anti-potato SBEII; starch branching enzyme; SBE A.
XX      Solanum tuberosum.
OS      Unidentified.
SS      Chimeric.
PN      EP118669-A2.
PD      25-JUL-2001.
PF      08-DEC-2000; 2000EP-00310997.
PR      17-DEC-1999; 99EP-00310188.
PA      (UNIL ) UNILEVER P.C.
PA      (UNIL ) UNILEVER NV.
PI      Franken LGJ, Van Der Logt CPE, Jobling SA, Teh Y;
DR      WPI: 2001-427157/46.
DR      N-PSDB; AAD10059.
PT      Modifying a plant to produce an antibody useful for increasing pathogen
PT      resistance or to modulate metabolism comprises introducing a DNA sequence
PT      encoding a heavy chain immunoglobulin linked to a peptide that targets a
PT      cellular compartment.
XX      Example 13; Fig 26; 81pp; English.
XX      The present invention relates to a method for modifying a plant to
XX      produce an antibody or an active fragment or derivative, or a protein
XX      functional equivalent, in a cellular compartment comprising introducing a
XX      DNA sequence encoding a heavy chain immunoglobulin, where the DNA is
XX      linked to promoters and provided with an additional sequence encoding a
XX      peptide capable of targeting heavy chain immunoglobulin to a cellular
XX      compartment. The method is used for producing a heavy chain
XX      immunoglobulin or an active fragment or derivative, or a protein that is
XX      functionally equivalent for increasing the pathogen resistance in a plant
XX      or to modulate metabolism in a plant. Under some circumstances it may be
XX      desirable to retain the antibody product with the plant rather than
XX      extracting and isolating the product. In particular, edible selected
XX      preferences may be used in a method of passively immunising an animal, the
XX      antigenically human, against the antigen, e.g., pathogenic organisms. The
XX      present sequence is an anti-potato SBEII antibody (denoted Clone68) heavy
XX      chain variable domain (VH) attached to peptide linkers, myc and his6 tag.
XX      The potato SBEII is a starch branching enzyme also designated SBE A
XX      Sequence 149 AA.

```

	Best Local Similarity	72.4%; Pred. No. 2,1e-35;	
	Matches	97; Conservative	6; Mismatches 12; Indels 19; Gaps 2
QY	1 QVQLDSSGGGLVQAAGSTFLSCAASGSIFFSDLLMGWTRQAPGREAVNAITRGTTVA	60	
DY	3 QVQLDSGGGLVQAAGSLRLSCAASGSIFFRPHNGWRFOAPGOGRRLVALISAGGTWA	62	
DY	61 DSVKGFETSRDNANNTWYLQNMNLSKEPDFAVVYCINARSRNSYDRSMGDYWGQGYTVSS	120	
DY	63 DSVKGRFTISRNNAKNTIYLQMNSLKPEDTIVYTCTA-----GSYGQGCGQTVA	114	
QY	121 -----AHH 123		
DY	115 EPKTPKPAAAH 128		
RESULT 12			
ID	AAE10555 standard; peptide; 131 AA.		
XX	AAE10555;		
XX	AAE10555;		
DT	10-DEC-2001 (first entry)		
DE	HPL inhibiting VHH fragment, HPL #15 from llama species.		
KM	Llama antibody; camelid; anorectic; heavy chain variable domain; VHH;		
KM	human dietary enzyme inhibitor; medicament; human pancreatic lipase; HPL;		
KM	food; human gastric lipase; HGL; cosmetic control; body weight.		
OS	Lama sp.		
XX			
FH	Key	Location/Qualifiers	
FT	Region	/label= CDR1	
FT		/note= "Complementarity determining region 1"	
FT	Region	/label= CDR2	
FT		/note= "Complementarity determining region 2"	
FT	Region	/label= CDR3	
FT		/note= "Complementarity determining region 3"	
PN	EP1134231-A1.		
PD	19-SEP-2001.		
PF	20-FEB-2001; 2001EP-00200703.		
PR	14-MAR-2000; 2000EP-00200930.		
PA	(UNITL ) UNILEVER NV.		
PA	(UNITL ) UNILEVER PLC.		
I1	Bezemer S, Van De Burg M, De Haard JMW, Tareilus E,		
I1	WPI; 2001-572718/65.		
PT	New antibody or its fragments for inhibiting human dietary enzymes,		
PT	useful for cosmetic control of body weight of human beings, comprises		
PT	heavy chain variable domain derived from immunoglobulin naturally devoid		
PT	of light chains.		
PS	Example 2; Page 9; 37pp; English.		
CC	The patent discloses antibodies or their fragments comprising a heavy		
CC	chain variable domain (VHH) derived from an immunoglobulin naturally		
CC	devoid of light chains specific for inhibiting human dietary enzymes. The		
CC	antibodies of the invention are useful for the preparation of medicaments		
CC	or food for inhibiting the activity of one or more human dietary enzymes		
CC	especially human pancreatic lipase (HPL) or human gastric lipase (HGL)		
CC	which are useful for the cosmetic control of body weight of human beings.		
CC	The present peptide sequence is HPL inhibiting VHH fragment, HPL #15 from		

CC llama (camelid) species  
XX Sequence 131 AA:

Query Match 69.7%; Score 474; DB 4; Length 131;  
Best Local Similarity 72.7%; Pred. No. 3,1e-35;  
Matches 96; Conservative 13; Mismatches 15; Indels 8; Gaps 3;

QY 1 QVQLQDSGGGLVQAGSLRLSCAASGIFSSDLMGMWYRQAPGKERAVARIRGGTSTYA 60  
Db 1 QVQLQDSGGGLVQAGSLRLSCAASGIFSSDLMGMWYRQAPGKERAVARIRGGTSTYA 60

QY 61 DSVKGRFTISRDNKNTMYLQNNSLKPEDTAVYYCAAR-----RSNYDSMGDWGQGTQ 115  
Db 61 DSVKGRFTISRDNKNTMYLQNNSLKPEDTAVYYCAARHITTPAGSSNY--VYG-YWGHGTR 117

QY 116 VTVSSAHSEDP 127  
Db 118 VTVSSSEPKTEPK 129

## RESULT 13

AAB67777  
ID AAB67777 standard; protein; 119 AA.

XX AAB67777;

DT 11-SEP-2003 (revised)  
DT 11-JUN-2001 (first entry)

DE Amino acid sequence of a llama heavy chain variable domain.

KW Heavy chain variable region; llama; Malassezia furfur; dandruff;  
hair care.

XX Lama glama.

PN W0200119871-A2.

XX 22-MAR-2001.

XX 28-AUG-2000; 2000MO-EP008380.

XX 16-SEP-1999; 99EP-00307356.

XX (UNITL ) UNILEVER PLC.

PA (UNITL ) UNILEVER NV.  
PA (HIND-) HINDUSTAN LEVER LTD.

PI Frenken LGJ, Van Der Vaart JM;

DR WPI; 2001-257877/26.

XX Composition for use in targeting active agent, especially antimicrobial  
PT agent to scalp for treating, preventing dandruff, has active agent  
PT conjugated to antibody capable of binding specifically to Malassezia  
furfur.

XX Claim 13; Page 9; 50pp; English.

XX The present sequence represents the heavy chain variable region of an  
CC antibody isolated from llama, which was immunised with Malassezia furfur.  
CC M. furfur has been implicated in dandruff formation. The heavy chain  
CC variable region is conjugated to an active agent, and used to produce a  
CC composition for topical application, e.g. to the scalp. The topical  
CC composition, e.g. hair care products such as shampoos and conditioners,  
CC skin care lotions, shower gels, etc., is useful for targeting an active  
CC agent to a site at which M. furfur is present for the treatment and  
CC prevention of dandruff. (Updated on 11-SEP-2003 to standardise OS field)

XX Sequence 119 AA;

Query Match 68.8%; Score 467.5; DB 4; Length 119;

Best Local Similarity 78.3%; Pred. No. 1,1e-34;  
Matches 94; Conservative 8; Mismatches 17; Indels 1; Gaps 1;

QY 1 QVQLQDSGGGLVQAGSLRLSCAASGIFSSDLMGMWYRQAPGKERAVARIRGGTSTYA 60

Db 1 QVQLQDSGGGLVQAGSLRLSCAASGIFSSDLMGMWYRQAPGKERAVARIRGGTSTYA 60

QY 61 DSVKGRFTISRDNKNTMYLQNNSLKPEDTAVYYCAARSDYRSGDWGQGTQYTVSS 120  
Db 61 DSVKGRFTISRDNKNTMYLQNNSLKPEDTAVYYCAAEDEHRIGTVG-YWGGGTQYTVSS 119

## RESULT 14

AAB67778  
ID AAB67778 standard; protein; 119 AA.

XX AAB67778;

DT 11-SEP-2003 (revised)  
DT 11-JUN-2001 (first entry)

DE Amino acid sequence of a llama heavy chain variable domain.

KW Heavy chain variable region; llama; Malassezia furfur; dandruff;  
hair care.

XX Lama glama.

PN W0200119871-A2.

XX 22-MAR-2001.

XX 28-AUG-2000; 2000MO-EP008380.

XX 16-SEP-1999; 99EP-00307356.

XX (UNITL ) UNILEVER PLC.

PA (UNITL ) UNILEVER NV.  
PA (HIND-) HINDUSTAN LEVER LTD.

PI Frenken LGJ, Van Der Vaart JM;

DR WPI; 2001-257877/26.

XX Composition for use in targeting active agent, especially antimicrobial  
PT agent to scalp for treating, preventing dandruff, has active agent  
PT conjugated to antibody capable of binding specifically to Malassezia  
furfur.

XX Claim 13; Page 9; 50pp; English.

XX The present sequence represents the heavy chain variable region of an  
CC antibody isolated from llama, which was immunised with Malassezia furfur.  
CC M. furfur has been implicated in dandruff formation. The heavy chain  
CC variable region is conjugated to an active agent, and used to produce a  
CC composition for topical application, e.g. to the scalp. The topical  
CC composition, e.g. hair care products such as shampoos and conditioners,  
CC skin care lotions, shower gels, etc., is useful for targeting an active  
CC agent to a site at which M. furfur is present for the treatment and  
CC prevention of dandruff. (Updated on 11-SEP-2003 to standardise OS field)

XX Sequence 119 AA;

Query Match 68.5%; Score 465.5; DB 4; Length 119;  
Best Local Similarity 78.3%; Pred. No. 1,1e-34;  
Matches 94; Conservative 7; Mismatches 18; Indels 1; Gaps 1;

QY 1 QVQLQDSGGGLVQAGSLRLSCAASGIFSSDLMGMWYRQAPGKERAVARIRGGTSTYA 60

Db 1 QVQLQDSGGGLVQAGSLRLSCAASGIFSSDLMGMWYRQAPGKERAVARIRGGTSTYA 60

QY 61 DSVKGRFTISRDNKNTMYLQNNSLKPEDTAVYYCAARSDYRSGDWGQGTQYTVSS 120  
Db 61 DSVKGRFTISRDNKNTMYLQNNSLKPEDTAVYYCAAEDEHRIGTVG-YWGGGTQYTVSS 119



This Page Blank (upside)

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2004, 07:36:16 ; Search time 21.0648 Seconds  
(without alignments)  
593.639 Million cell updates/sec

Title: US-09-805-290A-19

Perfect score: 685  
Sequence: 1 QVQLQESGGGLVQAGGSLRL...MGQGLTVVSSSEPTKPPKPP 130

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0 %  
Maximum Match 100 %

Listing first 45 summaries

Database : PIR\_78:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	388	56.6	119	2	S31107 Ig heavy chain - h
2	386	56.4	140	2	S31588 Ig heavy chain V r
3	385	56.2	119	2	S31108 Ig heavy chain - h
4	384.5	56.1	120	2	S48798 Ig heavy chain V r
5	381	55.6	160	2	S05271 Ig heavy chain pre
6	378.5	55.3	147	2	S17780 Ig variable region
7	378	55.2	121	2	S15673 Ig heavy chain - h
8	378	55.2	123	2	S31114 Ig heavy chain - h
9	378	55.2	138	2	S31666 Ig heavy chain V r
10	376	54.9	119	2	D36005 Ig heavy chain V r
11	376	54.9	121	2	S26798 Ig heavy chain V r
12	375	54.7	119	2	C36005 Ig heavy chain V r
13	374	54.6	123	2	S26794 Ig heavy chain V r
14	372.5	54.4	116	2	S31110 Ig heavy chain - h
15	372.5	54.3	127	2	S38489 Ig heavy chain - h
16	371.5	54.2	114	2	S46391 Ig heavy chain V r
17	371.5	54.2	128	2	S48797 Ig heavy chain V r
18	371.5	54.2	140	2	S70442 Ig heavy chain pre
19	371.5	54.2	151	2	S60943 Ig heavy chain pre
20	371	54.2	125	2	S40531 Ig heavy chain V r
21	370.5	54.1	120	2	S31101 Ig heavy chain V-D
22	370	54.0	119	2	F36005 Ig heavy chain V r
23	370	54.0	120	2	S55336 Ig heavy chain V r
24	370	54.0	122	2	E27888 Ig heavy chain V r
25	369	53.9	140	2	S23657 Ig heavy chain pre
26	368.5	53.8	124	2	S20782 Ig heavy chain V r
27	368	53.7	135	2	S31598 Ig heavy chain V r
28	368	53.7	138	2	A30561 Ig heavy chain pre
29	368	53.7	139	2	S31678 Ig heavy chain V r

30	368	53.7	140	2	S31686	Ig heavy chain V r
31	367	53.6	117	2	S78486	Ig heavy chain V r
32	367	53.6	127	2	S19878	Ig heavy chain V r
33	367	53.6	132	2	S31603	Ig heavy chain V r
34	366.5	53.5	119	2	D27889	Ig heavy chain V r
35	366.5	53.5	122	2	E36005	Ig heavy chain V r
36	366.5	53.5	128	2	S26786	Ig heavy chain V r
37	366	53.4	117	2	S31109	Ig heavy chain V r
38	366	53.4	121	2	G36005	Ig heavy chain - h
39	365	53.3	134	2	S31699	Ig heavy chain V r
40	364	53.1	120	2	S36278	Ig heavy chain V r
41	363.5	53.1	122	2	S31117	Ig heavy chain - h
42	363	53.0	115	1	A2HUBU	Ig heavy chain V-I
43	363	53.0	120	2	S55337	Ig heavy chain V r
44	362	52.8	123	2	S31693	Ig heavy chain V r
45	362	52.8	134	2	S31679	Ig heavy chain V r

#### ALIGNMENTS

##### RESULT 1

S31107  
Ig heavy chain - human  
C:Species: Homo sapiens (man)  
C:Date: 02-Dec-1993 #sequence\_revision 26-May-1995 #text\_change 17-Mar-1999  
C:Accession: S31107  
R:Raapportet, F.M.; Timmers, E.; Kenter, M.J.H.; van Tol, M.J.D.; Vossen, J.M.; Schuurman, Eur. J. Immunol. 22, 247-251, 1992  
A:Title: Restricted utilization of germ-line V(H)3 genes and short diverse third compleme  
A:Reference number: S31104; MIDID:92111633; PMID:1730252  
A:Accession: S31107  
A:Status: Preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: mRNA  
A:Residues: 1-119 <PAA>  
A:Cross-references: EMBL:X62955  
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1991  
C:Superfamily: immunoglobulin V region; immunoglobulin homology  
C:Keywords: heterotetramer; immunoglobulin  
F:15-98/Domain: immunoglobulin homology <IMM>

Query Match 56.6%; Score 388; DB 2; Length 119;  
Best Local Similarity 67.2%; Pred. No. 8.8e-27;

Matches 82; Conservative 8; Mismatches 28; Indels 4; Gaps 2;

QY 1 QVQLQESGGGLVQAGGSLRLSCAASGSIHTMGWYRTPGKERDVATIQ-DGGSITNY 59  
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
Db 1 EVQLVSGGGLVQPGSLRLSCAASGFTFSYAMSVRQAPKGLWEVAISGSGSTYY 60  
QY 60 ADSVKRFTISRDNLTNTVYLQNNDIKPEDTAVYYCNADVPYRFSRYLEWGGTLTV 119  
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
Db 61 ADSVKRFTISRDNLTNTVYLQNNDIKPEDTAVYYCAKDP---GASYFDYWGQGLTV 117  
QY 120 SS 121  
Db 118 SS 119

RESULT 2  
S31588  
Ig heavy chain V region - human (fragment)  
C:Species: Homo sapiens (man)  
C:Date: 22-Nov-1993 #sequence\_revision 10-Nov-1995 #text\_change 23-Jul-1999  
C:Accession: S31588  
R:Cuisinier, A.M.; Gauthier, D.; Boubli, L.; Fougereau, M.; Tonnelie, C.  
Submitted to the EMBL Data Library, June 1992  
A:Description: Mechanisms that generate human immunoglobulin diversity operate from the  
A:Reference number: S31585  
A:Accession: S31588  
A:Status: Preliminary  
A:Molecule type: mRNA  
A:Residues: 1-140 <CUI>  
A:Cross-references: EMBL:Z14200; NID:g30957; PIDN:CAA78569.1; PID:g30958

C/Superfamily: immunoglobulin V region; immunoglobulin homology  
C/Keywords: heterotetramer; immunoglobulin  
F/34-117/Domain: immunoglobulin homology <IMM>

Query Match 56.4%; Score 386; DB 2; Length 140;  
Best Local Similarity 66.4%; Pred. No. 1.6e-26;  
Matches 81; Conservative 9; Mismatches 30; Indels 2; Gaps 2;

QY 1 QVQLQESGGGLVQAGSLRLSCAASGSGISHTMGWRQTPGKERDVATIQ-DGGSITNY 59  
D 20 EVQLLESGGGLVQPGGSLRLSCAASGFTFSYAMSVWRQAPGKLEWVSAISGSGSTYY 79  
QY 60 ADSVKGRFTISRNTLTNTVYLGQNDLKPEDTAVYYCNADVPYRYSRYLEWVGQGLTVTV 119  
D 80 ADSVKGRFTISRNTLTNTVYLGQNDLKPEDTAVYYCAKD-HDYSNITFDYWGQGLTVTV 138

QY 120 SS 121  
D 139 SS 140

## RESULT 3

S31108

Ig heavy chain - human

C/Species: Homo sapiens (man)

C/Date: 02-Dec-1993 #sequence\_revision 26-May-1995 #text\_change 17-Mar-1999

C/Accession: S31108

R/Kaplan, F.M.; Timmers, E.; Kenter, M.J.H.; van Tol, M.J.D.; Vossen, J.M.; Schuurman

Eur. J. Immunol. 22, 247-251, 1992

A/Title: Restricted utilization of germ-line V(H)3 genes and short diverse third complement

A/Reference number: S31104; PMID:92111633; PMID:1730252

A/Accession: S31108

A/Status: preliminary; nucleic acid sequence not shown; translation not shown

A/Molecule type: mRNA

A/Residues: 1-119 &lt;RAA&gt;

A/Cross-references: EMBL:X62956

A/Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1991

C/Superfamily: immunoglobulin V region; immunoglobulin homology

C/Keywords: heterotetramer; immunoglobulin

F/15-98/Domain: immunoglobulin homology &lt;IMM&gt;

Query Match 56.2%; Score 385; DB 2; Length 119;  
Best Local Similarity 68.0%; Pred. No. 1.6e-26;  
Matches 83; Conservative 7; Mismatches 28; Indels 4; Gaps 2;

QY 1 QVQLQESGGGLVQAGSLRLSCAASGSGISHTMGWRQTPGKERDVATIQ-DGGSITNY 59  
D 1 EVQLLESGGGLVQPGGSLRLSCAASGFTFSYAMSVWRQAPGKLEWVSAISGSGSTYY 60  
QY 60 ADSVKGRFTISRNTLTNTVYLGQNDLKPEDTAVYYCNADVPYRYSRYLEWVGQGLTVTV 119  
D 61 ADSVKGRFTISRNTLTNTVYLGQNDLKPEDTAVYYCAKD-HDYSNITFDYWGQGLTVTV 117

QY 120 SS 121  
D 118 SS 119

## RESULT 4

S48798

Ig heavy chain V region (anti-Sm, VH3/DXP4/JH4b) - human

C/Species: Homo sapiens (man)

C/Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 23-Jul-1999

C/Accession: S48798

R/Mahmoudi, M.; Edwards, J.; Cairns, E.; Bell, D.

submitted to the EMBL Data Library, October 1994

A/Description: Molecular characterization of natural human anti-Sm autoantibodies.

A/Reference number: S48797

A/Accession: S48798

A/Status: preliminary

A/Molecule type: mRNA

A/Residues: 1-120 &lt;MAH&gt;

A/Cross-references: EMBL:Z46382; NID:9562324; PIDN:CAA86521.1; PID:91340167

C/Superfamily: immunoglobulin V region; immunoglobulin homology  
C/Keywords: heterotetramer; immunoglobulin  
F/15-98/Domain: immunoglobulin homology <IMM>

Query Match 56.1%; Score 384.5; DB 2; Length 120;  
Best Local Similarity 68.0%; Pred. No. 1.8e-26;  
Matches 83; Conservative 9; Mismatches 27; Indels 3; Gaps 3;

QY 1 QVQLQESGGGLVQAGSLRLSCAASGSGISHTMGWRQTPGKERDVATIQ-DGGSITNY 59  
D 1 EVQLLESGGGLVQPGGSLRLSCAASGFTFSYAMSVWRQAPGKLEWVSAISGSGSTYY 60  
QY 60 ADSVKGRFTISRNTLTNTVYLGQNDLKPEDTAVYYCNADVPYRYSRYLEWVGQGLTVTV 119  
D 61 ADSVKGRFTISRNTLTNTVYLGQNDLKPEDTAVYYCAKD-RGF-WGQYDYMVGQGLTVTV 118

QY 120 SS 121  
D 119 SS 120

## RESULT 5

S05271

Ig heavy chain precursor - human (fragment)

C/Species: Homo sapiens (man)

C/Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 16-Aug-1996

C/Accession: S05271; S04602

R/Kishimoto, T.

submitted to the EMBL Data Library, March 1989

A/Reference number: S05270

A/Accession: S05271

A/Molecule type: mRNA

A/Residues: 1-160 &lt;KIS1&gt;

A/Cross-references: EMBL:X14584

R/Kishimoto, T.; Okajima, H.; Okumoto, T.; Taniguchi, M.

Nucleic Acids Res. 17, 4385, 1989

A/Title: Nucleotide sequences of the cDNAs encoding the V-regions of H- and L-chains of

A/Reference number: S04601; PMID:89296497; PMID:250644

A/Accession: S04602

A/Molecule type: mRNA

A/Residues: 1-144 &lt;KIS2&gt;

A/Cross-references: EMBL:X14584

C/Superfamily: immunoglobulin V region; immunoglobulin homology

C/Keywords: heterotetramer; immunoglobulin

F/1-19/Domain: signal sequence #status predicted &lt;SIG&gt;

F/20-160/Product: Ig heavy chain (fragment) #status predicted &lt;MAT&gt;

F/34-117/Domain: immunoglobulin homology &lt;IMM&gt;

Query Match 55.6%; Score 381; DB 2; Length 160;  
Best Local Similarity 66.4%; Pred. No. 4.9e-26;  
Matches 83; Conservative 10; Mismatches 28; Indels 4; Gaps 3;

QY 1 QVQLQESGGGLVQAGSLRLSCAASGSGISHTMGWRQTPGKERDVATIQ-DGGSITNY 59  
D 20 EVQLLESGGGLVQPGGSLRLSCAASGFTFSYAMSVWRQAPGKLEWVSAISGSGSTYY 79  
QY 60 ADSVKGRFTISRNTLTNTVYLGQNDLKPEDTAVYYC-NADVPYRYSRY-LEWVGQGLTV 116  
D 80 ADSVKGRFTISRNTLTNTVYLGQNDLKPEDTAVYYCAKAVRGVIVSYYYGMDVWGQGLTV 139

QY 117 VTSS 121  
D 140 VTSS 144

## RESULT 6

I37780

Ig variable region (VDJ) (clone T20-11) - human (fragment)

C/Species: Homo sapiens (man)

C/Date: 16-Feb-1996 #sequence\_revision 16-Feb-1996 #text\_change 23-Jul-1999

C/Accession: I37780; S25474

R/Demaillon, C.; Chastagner, P.; Theze, J.; Zouali, M.

Proc. Natl. Acad. Sci. U.S.A. 91, 514-518, 1994

A/Title: Somatic diversification in the heavy chain variable region genes expressed by H  
A/Reference number: A26876; MUID:94119917; PMID:8290556  
A/Accession: I37780  
A/Status: Preliminary  
A/Molecule type: mRNA  
A/Residues: 1-147 <RES>  
A/Cross-references: EMBL:X67943; NID:g33578; PIDN:CAA48130.1; PID:g33579  
C/Superfamily: immunoglobulin V region; immunoglobulin homology  
F:28-111/Domain: immunoglobulin homology <IMM>

Query Match 55.2%; Score 378; DB 2; Length 147;  
Best Local Similarity 65.3%; Pred. No. 7.4e-26;  
Matches 81; Conservative 9; Mismatches 31; Indels 3; Gaps 2;

QY 1 OVQLQESGGGLVQAGSLRLSCAASGSIIGSIHTMGWYRQTPGERDVAITQ-DGGSITNY 59  
Db 14 EVQLVESGGGLVQPGGSLRLSCAASGFTFSITMSWVRQAPGKLEWVSAISGSGSTYY 73

QY 60 ADSVKGRFTISRDNLTNTVYLQMNLIKPEDTAVYYCNADVPRPRTSRV--LEWGGGTLV 117  
Db 74 ADSVKGRFTISRDNKNTLYLQMNSLRADTAVYYCAKDGRGWGLYYKGDVWGQGTIV 133

QY 118 TVSS 121  
Db 134 TVSS 137

RESULT 7  
155673  
Ig heavy chain - human (fragment)  
C/Species: Homo sapiens (man)  
C/Date: 02-Jul-1996 #revision 02-Jul-1996 #text\_change 21-Jan-2000  
C/Accession: I55673  
R/Knight, G.B.; Agnello, V.; Bonagura, V.; Barnes, J.L.; Panka, D.J.; Zhang, Q.X.  
J. Exp. Med. 178, 1903-1911, 1993  
A/Title: Human rheumatoid factor cross-idiotypes. IV. Studies on WA XId-positive IGM with  
elute from the 17.109 and G6 XId8.  
A/Reference number: I55673; MUID:94065558; PMID:8245772  
A/Accession: I55673  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: mRNA  
A/Residues: 1-121 <RES>  
A/Cross-references: GB:M87268; NID:g186197; PIDN:AC37536.1; PID:g186198  
C/Genetic8:  
A/Gene: GDB:IGHM  
A/Cross-references: GDB:120086; OMTM:147020  
A/Map position: 14q32.33-14q32.33  
C/Superfamily: immunoglobulin V region; immunoglobulin homology  
F:15-98/Domain: immunoglobulin homology <IMM>

Query Match 55.2%; Score 378; DB 2; Length 121;  
Best Local Similarity 66.4%; Pred. No. 6.6e-26;  
Matches 81; Conservative 9; Mismatches 30; Indels 2; Gaps 2;

QY 1 OVQLQESGGGLVQAGSLRLSCAASGSIIGSIHTMGWYRQTPGERDVAITQ-DGGSITNY 59  
Db 1 EVQLVESGGGLVQPGGSLRLSCAASGFTFSITMSWVRQAPGKLEWVSAISGSGSTYY 60

QY 60 ADSVKGRFTISRDNLTNTVYLQMNLIKPEDTAVYYCNADVPRPRTSRVLEWGGGTLV 119  
Db 61 ADSVKGRFTISRDNKNTLYLQMNSLRADTAVYYCAAPRHAGSPFY-DYWGQGTIV 119

QY 120 SS 121  
Db 120 SS 121

RESULT 8  
53114  
Ig heavy chain - human  
C/Species: Homo sapiens (man)  
C/Date: 02-Dec-1993 #sequence\_revision 26-May-1995 #text\_change 17-Mar-1999  
C/Accession: S3114

R/Raphorst, F.M.; Timmers, E.; Kenter, M.J.H.; van Tol, M.J.D.; Vossen, J.M.; Schuurman,  
Eur J. Immunol. 22, 247-251, 1992  
A/Title: Restricted utilization of germ-line VH3 genes and short diverse third compleme  
A/Reference number: S31104; MUID:9211633; PMID:1730252  
A/Accession: S31114  
A/Status: preliminary; nucleic acid sequence not shown; translation not shown  
A/Molecule type: mRNA  
A/Residues: 1-123 <RAA>  
A/Cross-references: EMBL:X62963  
A/Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1991  
C/Superfamily: immunoglobulin V region; immunoglobulin homology  
C/Keywords: heterotrimer; immunoglobulin  
F:15-98/Domain: immunoglobulin homology <IMM>

Query Match 55.2%; Score 378; DB 2; Length 123;  
Best Local Similarity 63.6%; Pred. No. 6.7e-26;  
Matches 82; Conservative 9; Mismatches 24; Indels 14; Gaps 3;

QY 1 OVQLQESGGGLVQAGSLRLSCAASGSIIGSIHTMGWYRQTPGERDVAITQ-DGGSITNY 59  
Db 1 EVQLVESGGGLVQPGGSLRLSCAASGFTFSITMSWVRQAPGKLEWVSAISGSGSTYY 60

QY 60 ADSVKGRFTISRDNLTNTVYLQMNLIKPEDTAVYYCNADVPRPRTSRV-----EWG 112  
Db 61 ADSVKGRFTISRDNKNTLYLQMNSLRADTAVYYC-----AKASLYRLEWLFDPYG 114

QY 113 QGTLTVSS 121  
Db 115 QGTLTVSS 123

RESULT 9  
531666  
Ig heavy chain V region - human (fragment)  
C/Species: Homo sapiens (man)  
C/Date: 22-Nov-1993 #sequence\_revision 10-Nov-1995 #text\_change 23-Jul-1999  
C/Accession: S31666  
R/Cuisinier, A.M.; Gauthier, L.; Boubli, L.; Fougereau, M.; Tonnelie, C.  
Submitted to the EMBL Data Library, June 1992  
A/Description: Mechanisms that generate human immunoglobulin diversity operate from the  
A/Reference number: S31585  
A/Accession: S31666  
A/Status: preliminary  
A/Molecule type: mRNA  
A/Residues: 1-138 <CUI>  
A/Cross-references: EMBL:Z14202; NID:g30963; PIDN:CAA78571.1; PID:g30964  
C/Superfamily: immunoglobulin V region; immunoglobulin homology  
C/Keywords: heterotrimer; immunoglobulin  
F:34-117/Domain: immunoglobulin homology <IMM>

Query Match 55.2%; Score 378; DB 2; Length 138;  
Best Local Similarity 65.6%; Pred. No. 7.6e-26;  
Matches 80; Conservative 10; Mismatches 28; Indels 4; Gaps 2;

QY 1 OVQLQESGGGLVQAGSLRLSCAASGSIIGSIHTMGWYRQTPGERDVAITQ-DGGSITNY 59  
Db 20 EVQLVESGGGLVQPGGSLRLSCAASGFTFSITMSWVRQAPGKLEWVSAISGSGSTYY 79

QY 60 ADSVKGRFTISRDNLTNTVYLQMNLIKPEDTAVYYCNADVPRPRTSRVLEWGGGTLV 119  
Db 80 ADSVKGRFTISRDNKNTLYLQMNSLRADTAVYYCAKARTGY--WYDLMGRGTLV 136

QY 120 SS 121  
Db 137 SS 138

RESULT 10  
D36005  
Ig heavy chain V region (M43) - human  
C/Species: Homo sapiens (man)  
C/Date: 21-Dec-1990 #sequence\_revision 21-Dec-1990 #text\_change 16-Dec-1998  
C/Accession: D36005

R:Schroeder Jr., H.W.; Wang, J.Y.  
Proc. Natl. Acad. Sci. U.S.A. 87, 6146-6150, 1990  
A>Title: Preferential utilization of conserved immunoglobulin heavy chain variable gene  
A/Accession: D36005  
A/Status: preliminary  
A/Molecule type: mRNA  
A/Residues: 1-119 <SCH>  
A/Cross-references: GB:M34024  
C/Genetics: GDB:IGH@; TGHDI1  
A/Genes: GDB:IGH@; TGHDI1  
A/Cross-references: GDB:118731; OMIM:146910  
A/Map position: 14q32.33-14q32.33  
C/Superfamily: immunoglobulin V region; immunoglobulin homology  
C/Keywords: heterotrimer; immunoglobulin  
F:15-98/Domain: immunoglobulin homology <IMW>

Query Match 54.9%; Score 376; DB 2; Length 119;  
Best Local Similarity 64.8%; Pred. No. 9.7e-26;  
Matches 79; Conservative 10; Mismatches 29; Indels 4; Gaps 2;

QY 1 QVQLQESGGGLVQAGSLRLSCAASGSGISHTMGYRQTPGKERDVATIQ-DGSGTNY 59  
Db 1 EVQLVESGGGLVQPGGSLRLSCAASGFTFSYAMSVWRQAPGKGLWWSAISGSGSTYY 60  
QY 60 ADSYKGRFTISRDNTLTNTVYLQNNDLKPEDTAVYYCNADVPRYRTSRYLEWVGQGLTVV 119  
Db 61 ADSYKGRFTISRDNTLTNTVYLQNNSLRAEDTAVYYC---AKDMMNDMPWGQGLTVV 117  
QY 120 SS 121  
Db 118 SS 119

## RESULT 11

S26798  
Ig heavy chain V region - human  
C/Species: Homo sapiens (man)  
C/Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 20-Jun-2000  
C/Accession: S26798  
R/Mortari, F.; Newton, J.A.; Wang, J.Y.; Schroeder Jr., H.W.  
Eur. J. Immunol. 22, 241-245, 1992  
A>Title: The human cord blood antibody repertoire. Frequent usage of the V(H)7 gene fami  
A/Reference number: S26786; MUID:92111632; PMID:1130251  
A/Accession: S26798  
A/Status: preliminary  
A/Molecule type: mRNA  
A/Residues: 1-121 <MOR>  
A/Cross-references: EMBL:X61015; NID:G32795; PID:CAA43349.1; PID:G1335126  
C/Superfamily: immunoglobulin V region; immunoglobulin homology  
C/Keywords: heterotrimer; immunoglobulin  
F:15-98/Domain: immunoglobulin homology <IMW>

Query Match 54.9%; Score 376; DB 2; Length 121;  
Best Local Similarity 66.4%; Pred. No. 9.9e-26;  
Matches 81; Conservative 10; Mismatches 29; Indels 2; Gaps 2;

QY 1 QVQLQESGGGLVQAGSLRLSCAASGSGISHTMGYRQTPGKERDVATIQDGS-TRY 59  
Db 1 EVQLVESGGGLVQPGGSLRLSCAASGFTFSYAMSVWRQAPGKGLWWSAISKGGITNY 60  
QY 60 ADSYKGRFTISRDNTLTNTVYLQNNDLKPEDTAVYYCNADVPRYRTSRYLEWVGQGLTVV 119  
Db 61 ADSYKGRFTISRDNTLTNTVYLQNNSLRAEDTAVYYCVRALGRY-SYGYLDYWGQGLTVV 119  
QY 120 SS 121  
Db 120 SS 121

## RESULT 12

S31110  
Ig heavy chain V region (30p1) - human  
C/Species: Homo sapiens (man)

C/Species: Homo sapiens (man)  
C/Date: 21-Dec-1990 #sequence\_revision 21-Dec-1990 #text\_change 16-Aug-1996  
C/Accession: C36005  
R:Schroeder Jr., H.W.; Wang, J.Y.  
Proc. Natl. Acad. Sci. U.S.A. 87, 6146-6150, 1990  
A>Title: Preferential utilization of conserved immunoglobulin heavy chain variable gene  
A/Reference number: A36005; MUID:90349571; PMID:2117273  
A/Status: preliminary  
A/Molecule type: mRNA  
A/Residues: 1-119 <SCH>  
A/Cross-references: GB:M18513  
C/Superfamily: immunoglobulin V region; immunoglobulin homology  
C/Keywords: heterotrimer; immunoglobulin  
F:15-98/Domain: immunoglobulin homology <IMW>

Query Match 54.7%; Score 375; DB 2; Length 119;  
Best Local Similarity 65.6%; Pred. No. 1.2e-25;  
Matches 80; Conservative 8; Mismatches 30; Indels 4; Gaps 2;

QY 1 QVQLQESGGGLVQAGSLRLSCAASGSGISHTMGYRQTPGKERDVATIQ-DGSGTNY 59  
Db 1 EVQLVESGGGLVQPGGSLRLSCAASGFTFSYAMSVWRQAPGKGLWWSAISGSGSTYY 60  
QY 60 ADSYKGRFTISRDNTLTNTVYLQNNDLKPEDTAVYYCNADVPRYRTSRYLEWVGQGLTVV 119  
Db 61 ADSYKGRFTISRDNTLTNTVYLQNNSLRAEDTAVYYCAKDA---GWSGFDYWGQGLTVV 117  
QY 120 SS 121  
Db 118 SS 119

## RESULT 13

S26794  
Ig heavy chain V region - human  
C/Species: Homo sapiens (man)  
C/Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 17-Mar-1999  
C/Accession: S26794  
R/Mortari, F.; Newton, J.A.; Wang, J.Y.; Schroeder Jr., H.W.  
Eur. J. Immunol. 22, 241-245, 1992  
A>Title: The human cord blood antibody repertoire. Frequent usage of the V(H)7 gene fami  
A/Reference number: S26786; MUID:92111632; PMID:1130251  
A/Accession: S26794  
A/Status: preliminary  
A/Molecule type: mRNA  
A/Residues: 1-123 <MOR>  
A/Cross-references: EMBL:X61011  
C/Superfamily: immunoglobulin V region; immunoglobulin homology  
C/Keywords: heterotrimer; immunoglobulin  
F:15-98/Domain: immunoglobulin homology <IMW>

Query Match 54.6%; Score 374; DB 2; Length 123;  
Best Local Similarity 63.4%; Pred. No. 1.5e-25;  
Matches 78; Conservative 13; Mismatches 30; Indels 2; Gaps 2;

QY 1 QVQLQESGGGLVQAGSLRLSCAASGSGISHTMGYRQTPGKERDVATIQDGS-TRY 59  
Db 1 EVQLVESGGGLVQPGGSLRLSCAASGFTFSYAMSVWRQAPGKGLWWSAISGSGSTYY 60  
QY 60 ADSYKGRFTISRDNTLTNTVYLQNNDLKPEDTAVYYCNADVPRYRTSRYLEWVGQGLTVV 118  
Db 61 ADSYKGRFTISRDNTLTNTVYLQNNSLRAEDTAVYYCARSIKYDENYIGMDWVGQGLTVV 120  
QY 119 VSS 121  
Db 121 VSS 123

## RESULT 14

S31110  
Ig heavy chain - human  
C/Species: Homo sapiens (man)



C>Date: 02-Dec-1993 #sequence\_revision 26-May-1995 #text\_change 17-Mar-1999  
 C/Accession: S31110  
 R/Raaphorst, F.W.; Timmers, E.; Kenter, M.J.H.; van Tol, M.J.D.; Vossen, J.M.; Schuurman  
 Eur. J. Immunol. 22, 247-251, 1992  
 A/Title: Restricted utilization of germ-line V(H)3 genes and short diverse third complement  
 A/Reference number: S31104; MUID:92111633; PMID:1730252  
 A/Accession: S31110  
 A/Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A/Molecule type: mRNA  
 A/Residues: 1-116 <RAA>  
 A/Cross-references: EMBL:X62958  
 A/Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1991  
 C/Superfamily: immunoglobulin V region; immunoglobulin homology  
 C/Keywords: heterotetramer; immunoglobulin  
 F:15-97/Domain: immunoglobulin homology <IMM>

Query Match 54.4%; Score 372.5; DB 2; Length 116;  
 Best Local Similarity 63.6%; Pred. No. 1.9e-25;  
 Matches 77; Conservative 12; Mismatches 27; Indels 5; Gaps 1;

QY 1 QVQLQESGGGLVQAGGSRRLSCAASGSIQHTWGWRQTGKERDVATIQDGSSTYA 60  
 DB 1 EVQLVETGGGLVQPGGSRRLSCAASGFTVSSVYMSWVRQAPGKGLEWVSIVYSGSTYA 60

QY 61 DSVKGRFTISRDNLTNTYTLQWNDLKPEDTAVVYCNADVREPRTSRYLEWVGQGLVTYS 120  
 DB 61 DSVKGRFTISRDNLSKNTLYLRNNSLRADTAVVYCARDL-----GGGSDYWGQGLVTYS 115

QY 121 S 121  
 DB 116 S 116

## RESULT 15

S38489

Ig heavy chain - human (fragment)

C/Species: Homo sapiens (man)

C/Date: 06-Jan-1995 #sequence\_revision 06-Jan-1995 #text\_change 23-Jul-1999

C/Accession: S38489

R/Marks, J.D.; Ouwehand, W.H.; Bye, J.M.; Finern, R.; Gorlick, B.D.; Voak, D.; Thorpe, S

submitted to the EMBL Data Library, June 1993

A/Description: Human antibody fragments specific for human blood group antigens from a p

A/Reference number: S38488

A/Accession: S38489

A/Status: Preliminary

A/Molecule type: DNA

A/Cross-references: EMBL:Z23028; NID:9414025; PIDN:CAA80563.1; PID:9414026

C/Superfamily: immunoglobulin V region; immunoglobulin homology

C/Keywords: heterotetramer; immunoglobulin

F:15-98/Domain: immunoglobulin homology &lt;IMM&gt;

Query Match 54.3%; Score 372; DB 2; Length 127;  
 Best Local Similarity 62.5%; Pred. No. 2.3e-25;  
 Matches 80; Conservative 10; Mismatches 28; Indels 10; Gaps 3;

QY 1 QVQLQESGGGLVQAGGSRRLSCAASGSIQHTWGWRQTGKERDVATIQDGSSTNY 59  
 DB 1 QVQLVQSGGAVQPGGSRRLSCAASGFTFSYMSWVRQAPGKGLEWVSIVYSGSTY 60

QY 60 ADSVKGRTISRDNLTNTYTLQWNDLKPEDTAVVYCNADVREPRTSRY-----LEWV 111  
 DB 61 ADSVKGRTISRDNLSKNTLYLRNNSLRADTAVVYCNADVREPRTSRY-----LEWV 119

QY 112 GQGTLVTV 119  
 DB 120 GQGTLVTV 127

Search completed: October 5, 2004, 08:17:13  
 Job time : 21.0648 secs

This Page Blank (uspb)

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2004, 07:29:35 ; Search time 11.6358 seconds  
(without alignments)  
581.749 Million cell updates/sec

Title: US-09-805-290A-19

Sequence: 1 QVQLQESGGGLVQAGGSLRL...WGQGLTVVSSSEPKTPKPP 130

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SwissProt 42:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	363	53.0	115	1 HV3F_HUMAN	P01767 homo sapien
2	352.5	51.5	122	1 HV3G_HUMAN	P01768 homo sapien
3	347	50.7	119	1 HV3I_HUMAN	P01770 homo sapien
4	339.5	49.6	119	1 HV3N_HUMAN	P01775 homo sapien
5	337.5	49.3	126	1 HV3K_HUMAN	P01772 homo sapien
6	337	49.2	116	1 HV05_CARP	P19181 carassius a
7	335.5	48.0	114	1 HV3B_HUMAN	P01763 homo sapien
8	334.5	48.8	119	1 HV3M_HUMAN	P01774 homo sapien
9	331.5	48.4	120	1 HV3U_HUMAN	P01782 homo sapien
10	328.5	48.0	117	1 HV3C_HUMAN	P01764 homo sapien
11	328	47.9	115	1 HV3D_HUMAN	P01765 homo sapien
12	328	47.9	117	1 HV02_CANFA	P01784 canis fami
13	326.5	47.7	114	1 HV3T_HUMAN	P01781 homo sapien
14	325.5	47.5	116	1 HV3E_HUMAN	P01766 homo sapien
15	325	47.4	120	1 HV3P_HUMAN	P01777 homo sapien
16	324.5	47.4	119	1 HV3E_HUMAN	P01766 homo sapien
17	324.5	47.4	136	1 HV16_MOUSE	P01783 mus musculu
18	324	47.3	121	1 HV3J_HUMAN	P01771 homo sapien
19	323	47.2	119	1 HV3L_HUMAN	P01773 homo sapien
20	323.5	47.1	122	1 HV3A_HUMAN	P01762 homo sapien
21	322	47.0	115	1 HV33_MOUSE	P01802 mus musculu
22	321.5	46.9	117	1 HV30_HUMAN	P01776 homo sapien
23	320	46.7	115	1 HV32_MOUSE	P01801 mus musculu
24	318	46.4	113	1 HV30_MOUSE	P01799 mus musculu
25	317.5	46.4	119	1 HV40_MOUSE	P01810 mus musculu
26	317.5	46.4	122	1 HV20_MOUSE	P01789 mus musculu
27	316.5	46.2	98	1 HV57_MOUSE	P18528 mus musculu
28	316.5	46.2	119	1 HV37_MOUSE	P01807 mus musculu
29	316	46.1	123	1 HV24_MOUSE	P01793 mus musculu
30	315	46.0	97	1 HV56_MOUSE	P18527 mus musculu
31	315	46.0	123	1 HV23_MOUSE	P01792 mus musculu
32	314.5	45.9	119	1 HV38_MOUSE	P01808 mus musculu
33	313	45.7	113	1 HV27_MOUSE	P01796 mus musculu

34	313	45.7	123	1 HV25_MOUSE	P01794 mus musculu
35	312.5	45.6	117	1 HV54_MOUSE	P18525 mus musculu
36	312	45.5	123	1 HV22_MOUSE	P01791 mus musculu
37	311.5	45.5	111	1 HV35_MOUSE	P01804 mus musculu
38	311.5	45.5	122	1 HV21_MOUSE	P01790 mus musculu
39	311.5	45.5	122	1 HV3H_HUMAN	P01769 homo sapien
40	310.5	45.3	117	1 HV53_MOUSE	P18524 mus musculu
41	310.5	45.3	117	1 HV55_MOUSE	P18526 mus musculu
42	310	45.3	123	1 HV18_MOUSE	P01787 mus musculu
43	309	45.1	118	1 HV33_MOUSE	P01809 mus musculu
44	309	45.1	144	1 HV26_MOUSE	P01795 mus musculu
45	308	45.0	113	1 HV31_MOUSE	P01800 mus musculu

## ALIGNMENTS

RESULT 1  
ID HV3F\_HUMAN STANDARD; PRT; 115 AA.  
AC P01767;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Ig heavy chain V-II region SUT.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_Taxid=9606;  
RN [1]  
RP SEQUENCE.  
RX MEDLINE=78137069; PubMed=416441;  
RA Torano A., Putnam F.W.;  
RT "Complete amino acid sequence of the alpha 2 heavy chain of a human  
IG2 immunoglobulin of the A2m (2) allotype."  
RL Proc. Natl. Acad. Sci. U.S.A. 75:966-969(1978).  
CC -1- MISCELLANEOUS: THE SEQUENCE OF THE ALPHA-2, A2M(2) ALLOTYPES, C  
REGION OF THIS MYELOMA PROTEIN IS ALSO GIVEN.  
CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.  
DR PIR; A02050; AZHUBU.  
DR HSBP; P01789; IMCP.  
DR GO; GO:0005576; C:extracellular; NAS.  
DR GO; GO:0003823; F:antigen binding; NAS.  
DR GO; GO:0006955; P:immune response; NAS.  
DR InterPro; IPR007110; Ig-like.  
DR InterPro; IPR003596; Ig\_V.  
DR Pfam; PF00047; Ig\_1.  
DR SMART; SM00406; IGV\_1.  
DR PROSITE; PS00835; IG LIKE; 1.  
KW Immunoglobulin V region.  
FT DOMAIN 1 111 IG-LIKE.  
FT NON TER 115 115  
SQ SEQUENCE 115 AA; 12379 MW; 208876A7DF52DCF4 CRC64;

Query Match 53.0%; Score 363; DB 1; Length 115;  
Best Local Similarity 61.2%; Pred. No. 4.6e-28;

Matches 74; Conservative 16; Mismatches 25; Indels 6; Gaps 1;

QY 1 QVQLQESGGGLVQAGGSLRLSCAASGSIHRTMGYRQTPGKERPVAVATIDQGSNTYA 60  
DB 1 EELVETGGGLTQPGSLRLSCAASGFTVSBHSMWRAQFGALZWVAIRGGTTYA 60.  
QY 61 DSVKGFITSRDNTLVTLQNMNDLPEPTAVYCAVDVPRFRTSYLHVWGGLTVTS 120  
DB 61 DSVKGFITSRDNTLVTLQNMNDLPEPTAVYCAVDVPRFRTSYLHVWGGLTVTS 114  
QY 121 S 121  
DB 115 S 115  
RESULT 2  
HV3G\_HUMAN

ID HV3G\_HUMAN STANDARD; PRT; 122 AA.  
 AC P01768;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Ig heavy chain V-II region CAM.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 NCBI\_TaxID=9606;  
 [1]  
 RP MEDLINE=81013859; PubMed=6774332;  
 RX Lehman D.W., Putnam F.W.;  
 RT Amino acid sequence of the variable region of a human mu chain:  
 RT location of a possible JH segment."  
 RL Proc. Natl. Acad. Sci. U.S.A. 77:3239-3243(1980).  
 CC -1- MISCELLANEOUS: THIS MU CHAIN WAS ISOLATED FROM THE PLASMA OF A  
 CC PATIENT WITH MACROGLOBULINEMIA.  
 CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.  
 DR PIR; A02051; M3HUM.  
 DR HSSP; P01772; 2PB4.  
 DR GO; GO:0005576; C:extracellular; NAS.  
 DR GO; GO:0003823; P:antigen binding; NAS.  
 DR GO; GO:0006955; P:immune response; NAS.  
 DR InterPro; IPR007110; IG-like.  
 DR InterPro; IPR003596; IG\_V.  
 DR Pfam; PF00047; IG\_1.  
 DR SMART; SM00406; IGV; 1.  
 DR PROSITE; PSS0835; IG\_LIKE; 1.  
 DR Immunoglobulin V region; Pyrrolidone carboxylic acid.  
 FT DOMAIN 1 112  
 FT MOD RES 1 112  
 FT NON TER 1 112  
 FT MOD RES 1 112  
 FT NON TER 1 112  
 SQ SEQUENCE 122 AA; 13668 MW; A42D0F17D252FIC2 CRC64;  
 Query Match 51.5%; Score 352.5; DB 1; Length 122;  
 Best Local Similarity 59.8%; Pred. No. 5.1e-27;  
 Matches 73; Conservative 13; Mismatches 35; Indels 1; Gaps 1;  
 QY 1 QVQLVSGGGLVQAGSLRLSCAASGSIHTMGWYRTPGKERDVATIQ-DGGSTNY 59  
 DB 1 QVQLVSGGGLVQAGSLRLSCAASGSIHTMGWYRTPGKERDVATIQ-DGGSTNY 59  
 QY 60 ADSVKGFTISRDNLTNTVYLQWMDKPEDTAVYYCNADVPYRTSRYLEVWGQGLTVV 119  
 DB 61 ADSVKGFTISRDNLTNTVYLQWMDKPEDTAVYYCNADVPYRTSRYLEVWGQGLTVV 119  
 QY 120 SS 121  
 DB 121 SS 122  
 RESULT 3  
 HV3I\_HUMAN STANDARD; PRT; 119 AA.  
 AC P01770;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Ig heavy chain V-III region NIE.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 NCBI\_TaxID=9606;  
 [1]  
 RP MEDLINE=77070269; PubMed=826475;  
 RX Postenngl H., Hilschmann N.;  
 RT "The rule of antibody structure. The primary structure of a  
 RT monoclonal 1961 immunoglobulin (myeloma protein NIE). III. The  
 RT chymotryptic peptides of the H-chain, alignment of the tryptic  
 RT peptides and discussion of the complete structure."  
 RT

RL Hoppe-Seyler's Z. Physiol. Chem. 357:1571-1604(1976).  
 RN [2]  
 RP DISULFIDE BOND.  
 RX MEDLINE=77070267; PubMed=1002129;  
 RA Dreker L., Schwarz J., Reichel W., Hilschmann N.;  
 RT "Rule of antibody structure. The primary structure of a monoclonal  
 RT Ig1 immunoglobulin (myeloma protein NIE). I: Purification and  
 RT characterization of the protein, the L- and H-chains, the  
 RT cyranogen bromide cleavage products, and the disulfide bridges."  
 RL Hoppe-Seyler's Z. Physiol. Chem. 357:1515-1540(1976).  
 CC -1- MISCELLANEOUS: THIS CHAIN WAS ISOLATED FROM AN IG1 MYELOMA  
 CC PROTEIN.  
 CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.  
 DR PIR; A91668; G1HUM1.  
 DR HSSP; P01772; 2PB4.  
 DR GO; GO:0005576; C:extracellular; NAS.  
 DR GO; GO:0003823; P:antigen binding; NAS.  
 DR GO; GO:0006955; P:immune response; NAS.  
 DR InterPro; IPR007110; IG-like.  
 DR InterPro; IPR003596; IG\_V.  
 DR Pfam; PF00047; IG\_1.  
 DR SMART; SM00406; IGV; 1.  
 DR PROSITE; PSS0835; IG\_LIKE; 1.  
 DR Immunoglobulin V region; Pyrrolidone carboxylic acid.  
 FT DOMAIN 1 112  
 FT MOD RES 1 112  
 FT DISULFID 22 96  
 FT NON TER 119 119  
 FT NON TER 119 119  
 SQ SEQUENCE 119 AA; 13242 MW; C96335A655E165B CRC64;  
 Query Match 50.7%; Score 347; DB 1; Length 119;  
 Best Local Similarity 60.7%; Pred. No. 1.7e-26;  
 Matches 74; Conservative 16; Mismatches 28; Indels 4; Gaps 3;  
 QY 1 QVQLVSGGGLVQAGSLRLSCAASGSIHTMGWYRTPGKERDVATIQ-DGGSTNY 59  
 DB 1 QVQLVSGGGLVQAGSLRLSCAASGSIHTMGWYRTPGKERDVATIQ-DGGSTNY 59  
 QY 60 ADSVKGFTISRDNLTNTVYLQWMDKPEDTAVYYCNADVPYRTSRYLEVWGQGLTVV 119  
 DB 61 ADSVKGFTISRDNLTNTVYLQWMDKPEDTAVYYCNADVPYRTSRYLEVWGQGLTVV 117  
 QY 120 SS 121  
 DB 118 SS 119  
 RESULT 4  
 HV3N\_HUMAN STANDARD; PRT; 119 AA.  
 AC P01775;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Ig heavy chain V-III region LAY.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 NCBI\_TaxID=9606;  
 [1]  
 RP MEDLINE=75046755; PubMed=4139708;  
 RX Capra J.D., Kehoe J.M.;  
 RA "Structure of antibodies with shared idiotypic: the complete sequence  
 RT of the heavy chain variable regions of two immunoglobulin M  
 RT anti-gamma globulins."  
 RL Proc. Natl. Acad. Sci. U.S.A. 71:4032-4036(1974).  
 CC -1- MISCELLANEOUS: THIS CHAIN WAS ISOLATED FROM IGM WITH ANTI-GAMMA  
 CC GLOBULIN ACTIVITY.  
 CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.  
 DR PIR; A02058; M3HULY.  
 DR HSSP; P01772; 2PB4.  
 DR GO; GO:0005576; C:extracellular; NAS.

DR GO; GO:0003823; F:antigen binding; NAS.  
 DR CO; GO:0006955; F:immune response; NAS.  
 DR InterPro; IPR007110; IG-like.  
 DR InterPro; IPR003596; IG\_v.  
 DR Pfam; PF00047; IG; 1.  
 DR SMART; SM00406; IGV; 1.  
 DR PROSITE; PS00835; IG LIKE; 1.  
 DR Immunoglobulin V region.  
 FT DOMAIN 1 112 IG-LIKE.  
 FT NON\_TER 119 119  
 SQ SEQUENCE 119 AA; 12858 MW; D6338058794DCE5E CRC64;

Query Match 49.6%; Score 339.5; DB 1; Length 119;  
 Best Local Similarity 58.5%; Pred. No. 8.6e-26;  
 Matches 69; Conservative 14; Mismatches 34; Indels 1; Gaps 1;

QY 2 VQLQESGGGLVQAGSLRLSCAASGSIQSIHTMGWYRQTPGKERDVVA-TIODGGSTVYA 60  
 DB 2 VQLQESGGGLVQAGSLRLSCAASGFTFSASMSWVRQAPGKLEWVAMKXENGNDKHYA 61  
 QY 61 DSVKGRFTISRDNTLVYLNQNDLKPEPTAVYYCNADVPYRISRILEWGGGLVT 118  
 DB 62 DSVNGFTISRNDKNTLVYLNQNGLOAVSAIYCARDAGPVSDTFFAHWGGLVT 119

## RESULT 5

HYK\_HUMAN STANDARD; PRT; 126 AA.  
 ID HVK\_HUMAN  
 AC P01772;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Ig heavy chain V-II region KOL.  
 OS Homo sapiens (Human)  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE, AND DISULFIDE BONDS.  
 RA MEDLINE=83289131; PubMed=6684994;  
 RA Schmidt W.B., Jung H.-D., Palm W., Hilschmann N.;  
 RT "Three-dimensional structure determination of antibodies. Primary  
 structure of crystallized monoclonal immunoglobulin IgG1 KOL, I.";  
 RL Hope-Seyler's Z. Physiol. Chem. 364:713-747(1983).  
 RN [2]  
 RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).  
 RA MEDLINE=81072295; PubMed=7441755;  
 RA Marguier M., Deisenhofer J., Huber R., Palm W.;  
 RT "Crystallographic refinement and atomic models of the intact  
 immunoglobulin molecule KOL and its antigen-binding fragment at 3.0 A  
 and 1.0-A resolution.";  
 RL J. Mol. Biol. 141:369-391(1980).  
 CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.  
 DR PIR; A02055; G1HUKL.  
 DR PDB; 2FBA; 12-JUL-89.  
 DR PDB; 2IG2; 12-JUL-89.  
 DR GO; GO:0005576; C:extracellular; NAS.  
 DR GO; GO:0003823; F:antigen binding; NAS.  
 DR GO; GO:0006955; F:immune response; NAS.  
 DR InterPro; IPR007110; IG-like.  
 DR InterPro; IPR003596; IG\_v.  
 DR Pfam; PF00047; IGV; 1.  
 DR SMART; SM00406; IGV; 1.  
 DR PROSITE; PS00835; IG LIKE; 1.  
 DR Immunoglobulin V region; 3D-structure; Pyrrolidone carboxylic acid.  
 FT DOMAIN 1 112 IG-LIKE.  
 FT MOD\_RES 1 1 PYRROLIDONE CARBOXYLIC ACID.  
 FT DISULFID 205 96  
 FT STRAND 105 110  
 FT STRAND 3 7  
 FT TURN 11 12  
 FT TURN 14 15  
 FT STRAND 18 25

FT HELIX 29 31  
 FT STRAND 34 39  
 FT TURN 41 42  
 FT STRAND 45 51  
 FT TURN 53 54  
 FT STRAND 58 60  
 FT HELIX 62 64  
 FT STRAND 65 65  
 FT STRAND 66 67  
 FT TURN 68 73  
 FT STRAND 74 77  
 FT STRAND 78 83  
 FT HELIX 88 90  
 FT STRAND 92 99  
 FT STRAND 106 106  
 FT TURN 107 108  
 FT STRAND 109 109  
 FT STRAND 113 116  
 FT STRAND 120 124  
 FT NON\_TER 126 126  
 SQ SEQUENCE 126 AA; 13718 MW; E4D71B52B16F8776 CRC64;

Query Match 49.3%; Score 337.5; DB 1; Length 126;  
 Best Local Similarity 56.4%; Pred. No. 1.4e-25;  
 Matches 75; Conservative 13; Mismatches 26; Indels 19; Gaps 3;

QY 1 QVQLQESGGGLVQAGSLRLSCAASGSIQSIHTMGWYRQTPGKERDVVA-TIODGGSTVYA 59  
 DB 1 QVQLQESGGGLVQAGSLRLSCAASGFTFSASMSWVRQAPGKLEWVAMKXENGNDKHYA 60  
 QY 60 ADVKGRFTISRNDNTLVYLNQNDLKPEPTAVYYCNAD-----VPRYISRVL 108  
 DB 61 ADVKGRFTISRNDKNTLVYLNQNDLKPEPTAVYYCNAD-----VPRYISRVL 113  
 QY 109 EWMGGGLVTSS 121  
 DB 114 DVMGGGTPTVSS 126

## RESULT 6

HV05\_CARAV STANDARD; PRT; 116 AA.  
 ID HV05\_CARAV  
 AC P19181;  
 DT 01-NOV-1990 (Rel. 16, Created)  
 DT 01-NOV-1990 (Rel. 16, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE Ig heavy chain V region 5A precursor.  
 OS Carassius auratus (Goldfish).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;  
 OC Cyprinidae; Carassius.  
 NCBI\_TaxID=7957;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=88144476; PubMed=3125551;  
 RA Wilson M.R., Middleton D., Watt G.W.;  
 RT "Immunoglobulin heavy chain variable region gene evolution: structure  
 and family relationships of two genes and a pseudogene in a teleost  
 fish.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 85:1566-1570(1988).  
 DR HSP; P01772; 2FBA.  
 DR InterPro; IPR007110; IG-like.  
 DR InterPro; IPR003596; IG\_v.  
 DR Pfam; PF00047; IGV; 1.  
 DR SMART; SM00406; IGV; 1.  
 DR PROSITE; PS00835; IG LIKE; 1.  
 DR Immunoglobulin V region; Signal.  
 FT SIGNAL 1 19  
 FT CHAIN 20 116 IG HEAVY CHAIN V REGION 5A.  
 FT DOMAIN 20 49 FRAMEWORK-1.  
 FT DOMAIN 50 54 COMPLEMENTARITY-DETERMINING-1.  
 FT DOMAIN 55 68 FRAMEWORK-2.  
 FT DOMAIN 69 84 COMPLEMENTARITY-DETERMINING-2.

FT DOMAIN 85 116 FRAMEWORK-3.  
 FT DISULFID 41 114 BY SIMILARITY.  
 FT NON TER 116  
 SQ SEQUENCE 116 AA; 12808 MW; 9C2279E2DF199B12 CRC64;

Query Match 49.2%; Score 337; DB 1; Length 116;  
 Best Local Similarity 70.5%; Pred. No. 1.5e-25;  
 Matches 67; Conservative 8; Mismatches 20; Indels 0; Gaps 0;

QY 1 QVQLQESGGGLVQAGGSLRLSCAASGSIHTMGWYRQTPGKERDVVAITIDGGSTNYA 60  
 DB 20 EVQLVDSGGGLIQPGSLRLSCAASGFTVSSNVMWVRQPPFGKLEWVSIVYSGSTIYA 79  
 QY 61 DSVKGRFTISRDNLTNTVYLQANDLKPEPTAVYYC 95  
 DB 80 DSVKGRFTISRDNLSKNTLYLQNMSLRAEDTAVYYC 114

## RESULT 7

HV3U HUMAN STANDARD; PRT; 114 AA.

AC P01763;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Ig heavy chain V-II region WEA.  
 OS Homo sapiens (human)  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 NCBI\_TaxID=9606;  
 RX MEDLINE=83273707; PubMed=6410398;  
 RA Goni F., Frangione B.;  
 RT "Amino acid sequence of the Fv region of a human monoclonal IGM  
 RT (protein WEA) with antibody activity against 3',4'-pyruvylated  
 RT galactose in Klebsiella polysaccharides K30 and K33."  
 RL Proc. Natl. Acad. Sci. U.S.A. 80:4837-4841(1983).  
 CC -1- MISCELLANEOUS: THIS CHAIN WAS OBTAINED FROM A MONOCLONAL ANTIBODY  
 CC AGAINST 3',4'-PYRUVYLATED GALACTOSE AND ISOLATED FROM A PATIENT WITH  
 CC WAGDENSTROM'S MACROGLOBULINEMIA.  
 CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.  
 DR PIR; A02046; M3HUPM.  
 DR HSSP; P01772; 2FB4.  
 DR GO; GO:0005576; C:extracellular; NAS.  
 DR GO; GO:0003823; F:antigen binding; NAS.  
 DR GO; GO:0006955; P:immune response; NAS.  
 DR InterPro; IPR003596; IG\_V.  
 DR InterPro; IPR003596; IG\_V.  
 DR Pfam; PF00047; IG\_1.  
 DR SMART; SM00406; IGV; 1.  
 DR PROSITE; PS50835; IG\_LIKE; 1.  
 DR Immunoglobulin V region; Pyroliadone carboxylic acid.  
 KM Immunoglobulin V region; Pyroliadone carboxylic acid.  
 FT DOMAIN 1 112  
 FT MOD RES 1 1  
 FT NON TER 114 114  
 SQ SEQUENCE 114 AA; 12256 MW; D83294FB418A07B7 CRC64;

Query Match 49.0%; Score 335.5; DB 1; Length 114;  
 Best Local Similarity 59.0%; Pred. No. 2e-25;  
 Matches 72; Conservative 15; Mismatches 26; Indels 9; Gaps 2;

QY 1 QVQLQESGGGLVQAGGSLRLSCAASGSIHTMGWYRQTPGKERDVVAITIDGGSTNY 59  
 DB 1 QVQLVDSGGGLVQAGGSLRLSCAASGFTVSSNVMWVRQPPFGKLEWVSIVYSGSTIYA 60  
 QY 60 ADSVKGFTISRDNLTNTVYLQANDLKPEPTAVYYCNADVRFPRISRILEVWGQGLTVY 119  
 DB 61 ADSVKGFTISRDNLSKNTLYLQNMSLRAEDTAVYYC-----ARGWLLNNGQGLTVY 112  
 QY 120 SS 121  
 DB 113 SS 114

## RESULT 8

HV3U HUMAN STANDARD; PRT; 119 AA.

AC P01774;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Ig heavy chain V-II region POM.  
 OS Homo sapiens (human)  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 NCBI\_TaxID=9606;  
 RX MEDLINE=75046755; PubMed=4139708;  
 RA Capra J.D., Kenoe J.M.;  
 RT "Structure of antibodies with shared idiotype: the complete sequence  
 RT of the heavy chain variable regions of two immunoglobulin M  
 RT anti-gamma globulins."  
 RL Proc. Natl. Acad. Sci. U.S.A. 71:4032-4036(1974).  
 CC -1- MISCELLANEOUS: THIS CHAIN WAS ISOLATED FROM IGM WITH ANTI-GAMMA  
 CC GLOBULIN ACTIVITY.  
 CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.  
 DR PIR; A02057; M3HUPM.  
 DR HSSP; P01772; 2FB4.  
 DR GO; GO:0005576; C:extracellular; NAS.  
 DR GO; GO:0003823; F:antigen binding; NAS.  
 DR GO; GO:0006955; P:immune response; NAS.  
 DR InterPro; IPR007110; IG\_V.  
 DR InterPro; IPR003596; IG\_V.  
 DR Pfam; PF00047; IG\_1.  
 DR SMART; SM00406; IGV; 1.  
 DR PROSITE; PS50835; IG\_LIKE; 1.  
 KM Immunoglobulin V region.  
 FT DOMAIN 1 112  
 FT VARIANT 54 54  
 FT VARIANT 54 54  
 FT NON TER 119 119  
 SQ SEQUENCE 119 AA; 12953 MW; 2E018AF4DCB2610 CRC64;

Query Match 48.8%; Score 334.5; DB 1; Length 119;  
 Best Local Similarity 57.1%; Pred. No. 2.e-25;  
 Matches 68; Conservative 16; Mismatches 34; Indels 1; Gaps 1;  
 QY 1 QVQLQESGGGLVQAGGSLRLSCAASGSIHTMGWYRQTPGKERDVVA-TIDGGSTNY 59  
 DB 1 EVQLVDSGGGLVQAGGSLRLSCAASGFTVSSNVMWVRQPPFGKLEWVSIVYSGSTIYA 60  
 QY 60 ADSVKGFTISRDNLTNTVYLQANDLKPEPTAVYYCNADVRFPRISRILEVWGQGLTVY 118  
 DB 61 ADSVKGFTISRDNLSKNTLYLQNMSLRAEDTAVYYCARDAQFPVSPTFPAHYGQGLTVY 119

## RESULT 9

HV3U HUMAN STANDARD; PRT; 120 AA.

AC P01782;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Ig heavy chain V-II region DOB.  
 OS Homo sapiens (human)  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 NCBI\_TaxID=9606;  
 RX MEDLINE=80020921; PubMed=114209;  
 RA Steiner L.A., Garcia Pardo A., Margolies M.N.;  
 RT "Amino acid sequence of the heavy-chain variable region of the

```

RT Crystallizable human myeloma protein Dob.;
RL Biochemistry 18:4068-4080(1979).
RN (2)
RP CRYSTALLIZATION.
RX MEDLINE=80020920; PubMed=114208;
RA Steiner L.A.; Lopes A.D.;
RT "The crystallizable human myeloma protein Dob has a hinge-region
deletion."
RL Biochemistry 18:4054-4067(1979).
CC -1- MISCELLANEOUS: THIS GAMMA-1 MYELOMA PROTEIN HAS A DELETION IN THE
HINGE REGION. THERE ARE NO LIGHT-HEAVY OR INTER-HEAVY CHAIN
DISULFIDE BONDS.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.
DR PIR: A30431; ClHUB4.
DR HSSP: P01772; 2F84.
DR GO: GO:0005576; C:extracellular; NAS.
DR GO: GO:0003823; P:antigen binding; NAS.
DR GO: GO:0006955; P:immune response; NAS.
DR InterPro: IPR007110; IG-like.
DR InterPro: IPR003596; IG_v.
DR Pfam: PF00047; IG_1.
DR SMART: SM00406; IGV_1.
DR PROSITE: PS50835; IG_LIKE; 1.
KW Immunoglobulin V region.
FT DOMAIN 1 112 IG-LIKE.
FT NON TER 120 120
SQ SEQUENCE 120 AA; 13440 MW; 880DD8307C4B2627 CRC64;

Query Match 48.4%; Score 331.5; DB 1; Length 120;
Best Local Similarity 60.2%; Pred. No. 5,1e-25;
Matches 74; Conservative 11; Mismatches 33; Indels 5; Gaps 3;

OY 1 QVQLQESGGGLVQAGGSRIRSCAASGSIQHTMGWYRTPGKERDVATIQ-DGGSTNY 59
DB 1 EVQLVDSGGDLVQPSRLRSCAASGFNFHEVNWMLQCGKGFENYVSTITWNGGSVLY 60

OY 60 ADVSKGRFTISRDNLTNTVYLQMNDLKPEDTAVVYCNADVPPY-RTSPYLEWVGQGLVLT 118
DB 61 ADVSKGRFASISDNQKTLVYLQNLIRPEDTAFYCC--AKGYINQWGWDSGGTLVT 117

OY 119 VSS 121
DB 118 VSS 120

RESULT 10
H3V3C_HUMAN STANDARD; PRT; 117 AA.
ID HV3C_HUMAN
AC P01764;
RT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DE 10-OCT-2003 (Rel. 42, Last annotation update)
DE IG heavy chain V-II region VH26 precursor.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=81101090; PubMed=6450418;
RA Mathysens G.; Rabbits T.H.;
RT "Structure and multiplicity of genes for the human immunoglobulin
heavy chain variable region."
RL Proc. Natl. Acad. Sci. U.S.A. 77:6561-6565(1980).
CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

```

```

CC -----
DR EMBL: J00236; AAAS3516.1; -
DR EMBL: M35415; AAAS8735.1; -
DR PIR: A02047; H3H26.
DR PDB: 1HOU; 23-DEC-99.
DR Genew: HGNC:5545; IGHV@.
DR GO: GO:0005576; C:extracellular; NAS.
DR GO: GO:0003823; P:antigen binding; NAS.
DR GO: GO:0006955; P:immune response; NAS.
DR InterPro: IPR007110; IG-like.
DR InterPro: IPR003596; IG_v.
DR Pfam: PF00047; IG_1.
DR SMART: SM00406; IGV_1.
DR PROSITE: PS50835; IG_LIKE; 1.
KW Immunoglobulin V region; Signal; 3D-structure.
FT SIGNAL 1 19
FT CHAIN 20 117 IG HEAVY CHAIN V-II REGION VH26.
FT DOMAIN 20 >117 IG-LIKE.
FT NON TER 117 117
SQ SEQUENCE 117 AA; 12582 MW; E826733F1A3CB0F1 CRC64;

Query Match 48.0%; Score 328.5; DB 1; Length 117;
Best Local Similarity 69.8%; Pred. No. 9,6e-25;
Matches 67; Conservative 7; Mismatches 21; Indels 1; Gaps 1;

OY 1 QVQLQESGGGLVQAGGSRIRSCAASGSIQHTMGWYRTPGKERDVATIQ-DGGSTNY 59
DB 20 EVQLVDSGGDLVQPSRLRSCAASGFNFHEVNWMLQCGKGFENYVSTITWNGGSVLY 79

OY 60 ADVSKGRFTISRDNLTNTVYLQMNDLKPEDTAVVYCNADVPPY-RTSPYLEWVGQGLVLT 118
DB 80 GDSVSKGRFTISRDNLTNTVYLQMNDLKPEDTAVVYCNADVPPY-RTSPYLEWVGQGLVLT 115

RESULT 11
H3V3C_HUMAN STANDARD; PRT; 115 AA.
ID HV3C_HUMAN
AC P01765;
RT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DE 10-OCT-2003 (Rel. 42, Last annotation update)
DE IG heavy chain V-II region TIL.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE
RX MEDLINE=78005528; PubMed=409716;
RA Wang A.-C.; Wang I.Y.; Fudenberg H.H.;
RT "Immunoglobulin structure and genetics. Identity between variable
regions of a mu and a gamma2 chain."
RL J. Biol. Chem. 252:7192-7199(1977).
CC -1- MISCELLANEOUS: THE SEQUENCES OF THE V REGIONS OF THE HEAVY CHAINS
OF IGM AND IGG2 ISOLATED FROM A SINGLE PATIENT WITH BICLOAL
GAMMOPATHY ARE IDENTICAL. THEIR LIGHT CHAINS ARE APPARENTLY ALSO
IDENTICAL.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.
DR PIR: A02048; H3H26.
DR HSSP: P01772; 2F84.
DR GO: GO:0005576; C:extracellular; NAS.
DR GO: GO:0003823; P:antigen binding; NAS.
DR GO: GO:0006955; P:immune response; NAS.
DR InterPro: IPR007110; IG-like.
DR InterPro: IPR003596; IG_v.
DR Pfam: PF00047; IG_1.
DR SMART: SM00406; IGV_1.
DR PROSITE: PS50835; IG_LIKE; 1.
KW Immunoglobulin V region.
FT DOMAIN 1 108 IG-LIKE.
FT NON TER 115 115
SQ SEQUENCE 115 AA; 12356 MW; 4DC67D179F62326 CRC64;

```







This Page Blank (uspto)

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2004, 07:34:11 ; Search time 68.817 Seconds

(without alignments)  
596.081 Million cell updates/sec

Title: US-09-805-290A-19

Perfect score: 685

Sequence: 1 QVQIQQSGGGGLVQAGSLRL.....WGQGLTVTVGSSEPKTPKPP 130

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

SPTREMBL\_25:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mmc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_ricent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_ivirus:\*  
16: sp\_bacteriaph:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	384	56.1	597	4	Q96BB9
2	375.5	54.8	118	4	Q9UL72
3	373.5	54.5	487	11	Q99KA4
4	370.5	54.1	499	4	Q8NSK4
5	368	53.7	113	4	Q8UL90
6	359.5	52.5	613	4	Q8WUK1
7	357	52.1	121	4	Q9UL71
8	353	51.5	480	11	Q91XEL
9	351	51.2	486	11	Q91Z07
10	349	50.9	131	4	Q9UL88
11	349	50.9	147	4	Q9Y509
12	348	50.8	119	11	Q920E7
13	348	50.8	479	11	Q91WPS
14	348	50.8	573	4	Q8WU38
15	345.5	50.4	118	4	Q9UL91
16	345.5	50.4	122	4	Q9UL84

17	345	50.4	116	4	Q9UL93	Q9UL93 homo sapien
18	343.5	50.1	112	4	Q9HCC1	Q9HCC1 homo sapien
19	343	50.1	494	4	Q96K69	Q96K69 homo sapien
20	339	49.5	487	11	Q80Z17	Q80Z17 mus musculu
21	338.5	49.4	493	4	Q8NCL6	Q8NCL6 mus musculu
22	332.5	48.5	469	11	Q8R3V9	Q8R3V9 mus musculu
23	332.5	48.5	473	11	Q91Z05	Q91Z05 mus musculu
24	331.5	48.4	521	4	Q8N4Y9	Q8N4Y9 homo sapien
25	327.5	47.8	437	11	Q9R1A4	Q9R1A4 mus musculu
26	321.5	46.9	95	4	Q9ULB6	Q9ULB6 homo sapien
27	319.5	46.6	298	11	Q9QYF0	Q9QYF0 mus musculu
28	312.5	45.6	470	4	Q7Z5M1	Q7Z5M1 homo sapien
29	307.5	44.9	484	11	Q8VEA0	Q8VEA0 mus musculu
30	304.5	44.5	124	6	Q9N0W6	Q9N0W6 oryctolagus
31	302.5	44.2	124	6	Q9N0W4	Q9N0W4 oryctolagus
32	290.5	42.4	479	11	Q7TMR4	Q7TMR4 mus musculu
33	290.5	42.4	462	4	Q7Z351	Q7Z351 homo sapien
34	271.5	39.6	124	4	Q9UL92	Q9UL92 mus musculu
35	271.5	39.6	482	11	Q91X92	Q91X92 mus musculu
36	270	39.4	104	4	Q9UL87	Q9UL87 homo sapien
37	270	39.4	614	11	Q7TMT6	Q7TMT6 mus musculu
38	262.5	38.3	613	11	Q8VXC7	Q8VXC7 mus musculu
39	262	38.2	112	4	Q9UCP3	Q9UCP3 homo sapien
40	260.5	38.0	169	11	Q8VDC9	Q8VDC9 mus musculu
41	258	37.7	159	4	Q9GQSO	Q9GQSO homo sapien
42	257	37.5	125	4	Q9UL95	Q9UL95 homo sapien
43	256	37.4	119	4	Q9UL73	Q9UL73 homo sapien
44	255.5	37.3	147	11	Q9Z5S3	Q9Z5S3 mus musculu
45	254.5	37.2	121	11	Q99NG4	Q99NG4 mus musculu

## ALIGNMENTS

## RESULT 1

ID Q96BB9 PRELIMINARY; PRT; 597 AA.

AC Q96BB9; 01-DEC-2001 (TREMBLrel. 19, Created)

DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)

DT 01-DEC-2003 (TREMBLrel. 25, Last annotation update)

DE Hypothetical protein.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

OC NCBI\_Taxid=9606;

RN [1] SEQUENCE FROM N.A.

RC TISSUE=B-cell;

RA Strauberg R.;

RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; BC015760; AAH15760.1; -

DR InterPro; IPR007110; IG\_1ike.

DR InterPro; IPR003006; IG\_MHC.

DR InterPro; IPR003596; IG\_v.

DR Pfam; PR00047; IG\_5.

DR SMART; SM00406; IGV\_1.

DR PROSITE; PSS0835; IG\_LIKE; 5.

DR PROSITE; PSS00290; IG\_MHC; 3.

KW Hypothetical protein.

SQ SEQUENCE 597 AA; 65039 MW; 4FCA3AD8E8CE263D9 CRC64;

Query Match 56.1%; Score 384; DB 4; Length 597;

Best Local Similarity 66.4%; Pred. No. 2.2e-29;

Matches 83; Conservative 9; Mismatches 29; Indels 4; Gaps 3;

QY 1 QVQIQQSGGGGLVQAGSLRLSCAASGISGHTMGWYRQTPGKERDVATIQ-DGGSTNY 59

DB 20 EVQLLESGGGLVPGQSLRLSCAASGFSFSSYAMNWRQAPGKGLVMSISSGGSTTY 79

QY 60 ADSYKGRFTISRNITLNTVYLQWMDLKPEPTATYYCNADVPFRITS-RLT--EWWQGLT 116

DB 80 ADSYKGRFTISRNISDITVLYLQWMDLKPEPTATYYCAKDPKRGYSAGNTRDYWGQGLT 139

QY	117	VTVSS	121
Db	140	VTVSS	144

## RESULT 2

ID	Q9UL72;	PRELIMINARY;	PRT;	118 AA.
AD	Q9UL72;			
DT	01-MAY-2000 (TrEMBLrel. 13, Created)			
DT	01-MAY-2000 (TrEMBLrel. 13, Last sequence update)			
DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)			
DE	Mysin-reactive immunoglobulin heavy chain variable region (Fragment).			
OS	Homo sapiens (human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=98277139; PubMed=9614934;			
RA	Wu X., Liu B., Van der Merwe P.L., Kalis N.N., Berney S.M.,			
RA	Young D.C.;			
RT	"Mysin-reactive autoantibodies in rheumatic carditis and normal			
RT	fevus.";			
RL	Clin. Immunol. Immunopathol. 87:184-192(1998).			
RL	EMBL; AF035042; AAD56578.1; -.			
DR	PIR; S21205; S21205.			
DR	HSSP; P01772; 2PB4.			
DR	InterPro; IPR007110; Ig-like.			
DR	InterPro; IPR005586; Ig_v.			
DR	Pfam; PF00047; Ig; 1.			
DR	SMART; SMO0406; IGV; 1.			
DR	PROSITE; PSS0835; IG_LIKE; 1.			
FT	NON_TER 1			
FT	NON_TER 118			
SQ	SEQUENCE 118 AA; 12872 MW; 64D1A5944BD25CCA CRC64;			

### RESULT 3

ID	Q99K44	PRELIMINARY;	PRT; 487 AA.
AC	Q99K44		
DT	01-JUN-2001 (TrEMBLrel. 17, Created)		
DT	01-JUN-2001 (TrEMBLrel. 17, Last sequence update)		
DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)		
DE	Hypothetical protein.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus		
OX	NCBI_TaxID=10090;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	Strausberg R;		
RL	Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.		
QR	EMBL; BC004786; AA04786.1; -		

## RESULT 4

DR HSSP; P01810; 2FBF.  
DR InterPro; IPR007110; IG-like.  
DR InterPro; IPR003006; IG\_MHC.  
DR InterPro; IPR003596; IG\_v.  
DR Pfam; PF00047; IG\_4.  
DR SMART; SM00406; IGV; 1.  
DR PROSITE; PS00835; IG\_LIKE; 4.  
DR PROSITE; PS00290; IG\_MHC; 2.  
DR Hypothetical protein  
KM SEQUENCE 487 AA; 52554 MW; 7DC8E96DB333077B CRC64;

## RESULT 4

ID	GENSK4	PRELIMINARY;	PRT;	499 AA.
AC	06NSK4			
DT	01-OCT-2002 (TREMBLrel. 22, Created)			
DT	01-OCT-2002 (TREMBLrel. 22, Last sequence update)			
DT	01-OCT-2003 (TREMBLrel. 25, Last annotation update)			
DE	Hypothetical protein.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
OX	NCBI_TaxId=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	T1TSUB=Blood;			
RA	Strausberg R.;			
RL	Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; BC032249; AAH32249.1; -.			
DR	InterPro; IPR003599; I9_.			
DR	InterPro; IPR007110; I9_1-like.			
DR	InterPro; IPR003597; I9_C1.			
DR	InterPro; IPR003006; I9_MHC.			
DR	InterPro; IPR003596; I9_V.			
DR	Pfam; PR00047; I9_4.			
DR	SMART; SM00409; I9_4.			
DR	SMART; SM00407; I9C1.2.			
DR	SMART; SM00406; I9V.1.			
DR	PROSITE; PS00835; I9_LIKE; 4.			
DR	PROSITE; PS00290; I9_MHC; 1.			
KW	Hypothetical protein.			
SC	SEQUENCE 499 AA; 53376 MW; 93A5C89562054F32 CRC64;			
Query Match	54.1%;	Score 370.5;	DB 4;	Length 499;
Best Local Similarity	57.7%;	Pred. No.3.8e-28;		
Matches	79;	Conservative 16;	Mismatches 35;	Indels 7;
				Gaps 3
QY	1 QVQLQSGGGGLVQAGGSLRLSCAASGSGISGHTNGMYRQTPGKERDVAITQ-DGGSTNY 59			
DB	20 EVQLVDSGGGVVPGPGSGIRLSCATSGFTFDDSGASWYRQAPGGLDWVSINNGSTNY 79			
QY	60 ADSVKKRFITISPDNTLNTVYIQMDLKEPDIAVYYCAADVRPYRT-----SRYLEWAGQ 114			
DB	80 ADSVKKRFITISDNKSLYIQNMSLRVEDALTYCARDPYTKYCGSSGLGYIMDVWYKSG 139			
QY	115 TLTVVSS-EPKTPKQFP 130			

```

Db          140 TTVTVSSASPTSPKVPF 156

RESULT 5
Q9UL90      PRELIMINARY;      PRT;      113 AA.
AC          Q9UL90;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Myosin-reactive immunoglobulin heavy chain variable region
   (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
CX NCBI_TaxId=9606;

RN
RP SEQUENCE FROM N.A.
RX MEDLINE=98277139; PubMed=9614934;
RA Wu X., Liu B., Van der Merwe P.L., Kalis N.N., Berney S.M.,
RA Young D.C.;
RT "Myosin-reactive autoantibodies in rheumatic carditis and normal
   fetus.";
RL Clin. Immunol. Immunopathol. 87:184-192(1998).
DR EMBL; AF035024; AAD56260.1; -.
DR HSSP; P01772; 2PB4.
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003596; IG_v.
DR Pfam; PF00047; Ig_1.
DR SMART; SM00406; IGV_1.
DR PROSITE; PS50835; IG-LIKE; 1.
FT NON TER      1
FT NON TER      1
SQ SEQUENCE      113 AA; 12437 MW; ED57PDD19086D07F CRC64;

Query Match      53.7%; Score 368; DB 4; Length 113;
Best Local Similarity 64.8%; Pred. No. 9.8e-29;
Matches 79; Conservative 9; Mismatches 24; Indels 10; Gaps 2

QY      1 QVQLDSGGIVQAGSRLRSCAASGSIIGSIHTGMVQTGPKERDVATIQ-DGSGTNY 59
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1 EVQLVESGGGVQPGQSLRLSCAASGTFPSSGCMHWVQAPEGKGLHWAFIRYDSNKKY 60
QY      60 ADSVAGRTTISBDNTLNTVYIQMDLKPEDTAVYYCNADVPRYRTSYLEWVGGLTVY 119
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      61 ADSVAGRTTISRDNSKNTLYIQMNSLRADPTAVYYCARD-----LNTWGGGLTVY 111
QY      120 SS 121
      : :
Db      112 SS 113

RESULT 6
8MWU1      PRELIMINARY;      PRT;      613 AA.
AC          8MWU1;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
CX NCBI_TaxId=9606;

RN
RP SEQUENCE FROM N.A.
RX TISUE=Tonsil;
RA Strausberg R.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC020240; AAH20240.1; -.
DR PIR; P10120; P10120.
DR PIR; S15590; S15590.

```

Query Match	Best Local Similarity	Score	DB 4	Length	613
Matches 78; Conservative 11; Mismatches 26; Indels 11; Gaps 3	52.5%; Pred. 61.3%;	359.5;	DB 4;	Length 613;	
QY 1 QVQLVESGGGVQVQGGGSLRLSCAASGSGISHTMGWYKQTPGKERDVAATIQ-DGSGTNY 59	20 QVQLVESGGGVQVQGGGSLRLSCAASGTFSSFGYGHWRQAKGKLEWAVLSYDSGNKY 79	60 ADSYKGRFTISRDNLTNTVYIQMNDLKEPDYAVYYCNAD---VPRYTSRYLEWGGCT 115	80 ADSYKGRFTISRDNLTNTVYIQMNSLRAPEDYAVYYCAKDWSEGVETF-----DIWGCGT 133		
QY 116 LTVVSS 121	134 MVTVSS 139				
DB					
RESULT 7					
Q9UL71	PRELIMINARY;	PRT; 121 AA.			
AC Q9UL71.1					
DT 01-MAY-2000 (TIREMBL-rel. 13, Created)					
DT 01-MAY-2000 (TIREMBL-rel. 13, Last sequence update)					
DT 01-OCT-2003 (TIREMBL-rel. 25, Last annotation update)					
DE Myosin-reactive immunoglobulin heavy chain variable region (Fragment).					
DE Homo sapiens (Human).					
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.					
OX NCBI_TaxID=9606;					
RN [1]					
RP SEQUENCE FROM N.A.					
PX MEDLINE=9827139; PubMed=9614934;					
RA Wu X., Liu B., Van der Merwe P.L., Kalis N.N., Berney S.M.,					
RT Young D.C.;					
RT "Myosin-reactive autoantibodies in rheumatic carditis and normal fetus";					
RL Clin. Immunol. Immunopathol. 87:184-192(1998).					
DR EMBL; AF035043; AAD56279.1; -					
DR HSSP; P01772; 2PB4					
DR InterPro; IPR007110; IG_1like.					
DR InterPro; IPR003596; IG_v.					
DR Pfam; PF00047; IG; 1.					
DR SMART; SM00406; IGV; 1.					
DR PROSITE; PS00835; IG_LIKE; 5.					
DR PROSITE; PS00290; IG_MHC; 3.					
KW Hypothetical protein.					
SO SEQUENCE 613 AA; 67296 MW; 60CF75950671E315 CRC64;					
QY 52.5%; Score 359.5; DB 4; Length 613;					
Best Local Similarity 61.3%; Pred. No. 6e-27;					
Matches 78; Conservative 11; Mismatches 26; Indels 11; Gaps 3					
QY 1 QVQLVESGGGVQVQGGGSLRLSCAASGSGISHTMGWYKQTPGKERDVAATIQ-DGSGTNY 59					
DB 20 QVQLVESGGGVQVQGGGSLRLSCAASGTFSSFGYGHWRQAKGKLEWAVLSYDSGNKY 79					
QY 60 ADSYKGRFTISRDNLTNTVYIQMNDLKEPDYAVYYCNAD---VPRYTSRYLEWGGCT 115					
DB 80 ADSYKGRFTISRDNLTNTVYIQMNSLRAPEDYAVYYCAKDWSEGVETF-----DIWGCGT 133					
QY 116 LTVVSS 121					
DB 134 MVTVSS 139					
RESULT 7					
Q9UL71	PRELIMINARY;	PRT; 121 AA.			
AC Q9UL71.1					
DT 01-MAY-2000 (TIREMBL-rel. 13, Created)					
DT 01-MAY-2000 (TIREMBL-rel. 13, Last sequence update)					
DT 01-OCT-2003 (TIREMBL-rel. 25, Last annotation update)					
DE Myosin-reactive immunoglobulin heavy chain variable region (Fragment).					
DE Homo sapiens (Human).					
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.					
OX NCBI_TaxID=9606;					
RN [1]					
RP SEQUENCE FROM N.A.					
PX MEDLINE=9827139; PubMed=9614934;					
RA Wu X., Liu B., Van der Merwe P.L., Kalis N.N., Berney S.M.,					
RT Young D.C.;					
RT "Myosin-reactive autoantibodies in rheumatic carditis and normal fetus";					
RL Clin. Immunol. Immunopathol. 87:184-192(1998).					
DR EMBL; AF035043; AAD56279.1; -					
DR HSSP; P01772; 2PB4					
DR InterPro; IPR007110; IG_1like.					
DR InterPro; IPR003596; IG_v.					
DR Pfam; PF00047; IG; 1.					
DR SMART; SM00406; IGV; 1.					
DR PROSITE; PS00835; IG_LIKE; 1.					
DR PROSITE; PS00290; IG_MHC; 3.					
KW Hypothetical protein.					
SO SEQUENCE 613 AA; 67296 MW; 60CF75950671E315 CRC64;					
QY 52.5%; Score 359.5; DB 4; Length 613;					
Best Local Similarity 61.3%; Pred. No. 1.3e-27;					
Matches 76; Conservative 15; Mismatches 27; Indels 6; Gaps 3					
QY 1 QVQLVESGGGVQVQGGGSLRLSCAASGSGISHTMGWYKQTPGKERDVAATIQ-DGSGTNY 59					
DB 1 EVQLVESGGGVQVQGGGSLRLSCAASGTFPDGYAHWYQAKGKLEWVSLISGDSGITY 60					
QY 60 ADSYKGRFTISRDNLTNTVYIQMNDLKEPDYAVYYCNADVAPYRTSR--LEWGGCTLV 117					
DB 61 ADSYKGRFTISRDNLTNTVYIQMNSLRAPEDYAVYYCAKDWSEGVETF-----DIWGCGT 117					

Db 118 TVSS 121

## RESULT 8

Q91XEL PRELIMINARY; PRT; 480 AA.

AC Q91XEL; 01-DEC-2001 (T-EMBLrel. 19, Created)

DT 01-DEC-2001 (T-EMBLrel. 19, Last sequence update)

DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)

DE Hypothetical protein (Fragment).

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI\_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Colon;

RA Strausberg R.;

RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL: BC010798; AAH10798.1; -

DR InterPro: IPR007110; IG\_1ike.

DR InterPro: IPR003006; IG\_MHC.

DR Pfam: PF00047; IG\_4.

DR SMART: SM00406; IGV\_1.

DR PROSITE: PS00835; IG\_LIKE; 4.

DR PROSITE: PS00290; IG\_MHC; 2.

KW Hypothetical protein.

FT NON\_TER

SQ SEQUENCE 480 AA; 51936 MW; 20B9234EEF2B41ED CRC64;

Query Match 51.5%; Score 353; DB 11; Length 480;

Best Local Similarity 62.6%; Pred. No. 1.9e-26;

Matches 77; Conservative 8; Mismatches 32; Indels 6; Gaps 2;

QY 2 VQLQESGGGLVQAGSLRLSCAASGSIQTHMGWYRQTPGKERVAVTIQDGG-STNYA 60

DB 20 VKLVEGGGLVPGGSLRLSCAASGTFPSNYSWVWVQTPERKLEWVAITISNGYATHP 79

QY 61 DSVKGRFTISRDNTLVNTLVYLNQNDLKPEDTAVYVCNADVPRTSRYLEWVGQGT 120

DB 80 DSVKGRFTISRDNTLVNTLVYLNQNDLKPEDTAVYVCNADVPRTSRYLEWVGQGT 134

QY 121 SEP 123

DB 135 SEP 137

RESULT 9

Q91207 PRELIMINARY; PRT; 486 AA.

AC Q91207; 01-DEC-2001 (T-EMBLrel. 19, Created)

DT 01-DEC-2001 (T-EMBLrel. 19, Last sequence update)

DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)

DE Hypothetical protein.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI\_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RA Strausberg R.;

RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL: BC010324; AAH10324.1; -

DR InterPro: IPR007110; IG\_1like.

DR InterPro: IPR003006; IG\_MHC.

DR InterPro: IPR003596; IG\_V.

DR Pfam: PF00047; IG\_4.

DR SMART: SM00406; IGV\_1.

DR PROSITE: PS00835; IG\_LIKE; 4.

DR PROSITE: PS00290; IG\_MHC; 2.

KW Hypothetical protein.

SQ SEQUENCE 486 AA; 52682 MW; 4FEF835125DA870B CRC64;

Query Match 51.2%; Score 351; DB 11; Length 486;

Best Local Similarity 55.9%; Pred. No. 3.1e-26;

Matches 71; Conservative 16; Mismatches 32; Indels 8; Gaps 2;

QY 1 QVQLQESGGGLVQAGSLRLSCAASGSIQTHMGWYRQTPGKERVAVTIQDGGSTNYA 60

DB 20 EVHLVSGGGLVPGGSLRLSCVSGFSTYSMSWVWVQTPERKLEWVAITISNGNTYTP 79

QY 61 DSVKGRFTISRDNTLVNTLVYLNQNDLKPEDTAVYVCNADVP-----YRTSRYLEWVGQGT 115

DB 80 DSVKGRFTISRDNTLVNTLVYLNQNDLKPEDTAVYVCNADVP-----VREPIPIYSGSYFDSWGQGT 136

QY 116 LTVSS 122

DB 137 TTVSS 143

## RESULT 10

Q9UL88 PRELIMINARY; PRT; 131 AA.

AC Q9UL88; 01-MAY-2000 (T-EMBLrel. 13, Created)

DT 01-MAY-2000 (T-EMBLrel. 13, Last sequence update)

DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)

DE Myosin-reactive immunoglobulin heavy chain variable region

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=98277139; PubMed=9614934;

RA Wu X., Liu B., Van der Merwe P.L., Kalis N.N., Berney S.M.,

RT "Myosin-reactive autoantibodies in rheumatic carditis and normal

RL Clin. Immunol. Immunopathol. 87:184-192(1998).

DR EMBL: AF035026; AAD56262.1; -

DR PIR: S21205; S21205.

DR HSP: P01810; 2FBJ.

DR InterPro: IPR007110; IG\_1like.

DR InterPro: IPR003596; IG\_V.

DR Pfam: PF00047; IGV\_1.

DR SMART: SM00406; IGV\_1.

DR PROSITE: PS00835; IG\_LIKE; 1.

FT NON\_TER

SQ SEQUENCE 131 AA; 14142 MW; 96E7D668E375DEA0 CRC64;

Query Match 50.2%; Score 349; DB 4; Length 131;

Best Local Similarity 58.6%; Pred. No. 9e-27;

Matches 78; Conservative 11; Mismatches 30; Indels 14; Gaps 3;

QY 1 QVQLQESGGGLVQAGSLRLSCAASGSIQTHMGWYRQTPGKERVAVTIQDGGST 57

DB 1 EVQLVSGGGLVPGGSLRLSCVSGFSTYSMSWVWVQTPERKLEWVAITISNGNTYTP 60

QY 58 NVADSVKGRFTISRDNTLVNTLVYLNQNDLKPEDTAVYVCNADVP-----RYPRTSRYL 108

DB 61 DYAPVYKGRFTISRDNTLVNTLVYLNQNDLKPEDTAVYVCNADVP-----RYPRTSRYL 118

QY 109 EYWGQGLTVSS 121

DB 119 EYWGQGLTVSS 131

## RESULT 11

Q9Y509 PRELIMINARY; PRT; 147 AA.

AC Q9Y509; 01-NOV-1999 (Tremblrel. 12, Created)  
 DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)  
 DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)  
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)  
 DE VH3 protein (Fragment).  
 GN VH3.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 CC NCBI\_TaxID=9606;  
 RX MEDLINE=96071149; PubMed=7475288;  
 RA Cao J., Vescio R.A., Rettig M.B., Hong C.H., Kim A., Lee J.C.,  
 RA Lichtenstein A.K., Berenson J.R.,  
 RT "A CD10-positive subset of malignant cells is identified in multiple  
 RT leukemia using PCR with patient-specific immunoglobulin gene primers."  
 RL Leukemia 9:1948-1953(1995).  
 DR EMBL; S80860; AAD14339.1; -.  
 DR HSSP; P01772; 2FB4.  
 DR GO; GO:0005887; C:integral to plasma membrane; NAS.  
 DR GO; GO:0016489; F:immunoglobulin receptor activity; NAS.  
 DR GO; GO:0016066; P:cellular defense response (sensu Vertebrata); NAS.  
 DR InterPro; IPRO07110; IG-like.  
 DR Pfam; PF00047; Ig; 1.  
 DR SMART; SM00406; IgV; 1.  
 DR PROSITE; PS50835; IG-LIKE; 1.  
 FT NON\_TER 1  
 SQ SEQUENCE 147 AA; 15768 MW; 8489FCAA7BC925C CRC64;  
 Query Match 50.8%; Score 349; DB 4; Length 147;  
 Best Local Similarity 58.8%; Pred. No. 1e-26;  
 Matches 80; Conservative 9; Mismatches 41; Indels 6; Gaps 3;  
 QY 1 QVQLQESGGGLVQAGSRLSCAASGSIQSIHTMGWRQTGKERDVATITQDGSSTN-Y 59  
 DB 1 QVHVLVESGGGVQVQPKSLRLSCASGFTFTSTYGMWVRQAPGKGLDWALISYDGSSTQY 60  
 QY 60 ADSVKGRTISRDNLTNTVYLQNMNDLKPEDTAVYYCNADVPRPYTSRY---LEWVGQGT 115  
 DB 61 AGSVKGRFTISRDNKNTLYLQMSLRVEDIATVYCAADGVPFVSQVYVAGIYWGQGT 120  
 QY 116 LVTWSS-EPKTPKPOP 130  
 DB 121 LVTWSSASTKGPSVFP 136  
 RESULT 12  
 Q920E7 PRELIMINARY; PRT; 119 AA.  
 ID Q920E7;  
 AC Q920E7;  
 DT 01-DEC-2001 (Tremblrel. 19, Created)  
 DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)  
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)  
 DE Pterin-mimicking anti-idiotypic heavy chain variable region  
 (Fragment).  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 CC NCBI\_TaxID=10090;  
 RX MEDLINE=96071149; PubMed=7475288;  
 RA Cao J., Vescio R.A., Rettig M.B., Hong C.H., Kim A., Lee J.C.,  
 RA Lichtenstein A.K., Berenson J.R.,  
 RT "A CD10-positive subset of malignant cells is identified in multiple  
 RT leukemia using PCR with patient-specific immunoglobulin gene primers."  
 RL Leukemia 9:1948-1953(1995).  
 DR EMBL; S80860; AAD14339.1; -.  
 DR HSSP; P01772; 2FB4.  
 DR GO; GO:0005887; C:integral to plasma membrane; NAS.  
 DR GO; GO:0016489; F:immunoglobulin receptor activity; NAS.  
 DR GO; GO:0016066; P:cellular defense response (sensu Vertebrata); NAS.  
 DR InterPro; IPRO07110; IG-like.  
 DR Pfam; PF00047; Ig; 1.  
 DR SMART; SM00406; IgV; 1.  
 DR PROSITE; PS50835; IG-LIKE; 1.  
 FT NON\_TER 1  
 SQ SEQUENCE 147 AA; 15768 MW; 8489FCAA7BC925C CRC64;  
 Query Match 50.8%; Score 349; DB 4; Length 147;  
 Best Local Similarity 58.8%; Pred. No. 1e-26;  
 Matches 80; Conservative 9; Mismatches 41; Indels 6; Gaps 3;  
 QY 1 QVQLQESGGGLVQAGSRLSCAASGSIQSIHTMGWRQTGKERDVATITQDGSSTN-Y 59  
 DB 1 QVHVLVESGGGVQVQPKSLRLSCASGFTFTSTYGMWVRQAPGKGLDWALISYDGSSTQY 60  
 QY 60 ADSVKGRTISRDNLTNTVYLQNMNDLKPEDTAVYYCNADVPRPYTSRY---LEWVGQGT 115  
 DB 61 AGSVKGRFTISRDNKNTLYLQMSLRVEDIATVYCAADGVPFVSQVYVAGIYWGQGT 120  
 QY 116 LVTWSS-EPKTPKPOP 130  
 DB 121 LVTWSSASTKGPSVFP 136

DR PROSITE; PS50835; IG-LIKE; 1.  
 FT NON\_TER 1  
 FT NON\_TER 119  
 SQ SEQUENCE 119 AA; 13025 MW; F6E90404381CA7C CRC64;  
 Query Match 50.8%; Score 348; DB 11; Length 119;  
 Best Local Similarity 61.6%; Pred. No. 9.9e-27;  
 Matches 77; Conservative 9; Mismatches 29; Indels 10; Gaps 3;  
 QY 1 QVQLQESGGGLVQAGSRLSCAASGSIQSIHTMGWRQTGKERDVATITQDGSSTN-Y 59  
 DB 1 EVQLVESGGGLVQPGGSLKSCAASGFTFTSTYGMWVRQAPGKGLDWALISYDGSSTQY 60  
 QY 60 ADSVKGRTISRDNLTNTVYLQNMNDLKPEDTAVYYCNADVPRPYTSRY---LEWVGQGT 116  
 DB 61 ADSVKGRTISRDNKNTLYLQMSLRVEDIATVYCAADGVPFVSQVYVAGIYWGQGT 114  
 QY 117 VTYSS 121  
 DB 115 VTYSSA 119  
 RESULT 13  
 Q91WP5 PRELIMINARY; PRT; 479 AA.  
 ID Q91WP5;  
 AC Q91WP5;  
 DT 01-DEC-2001 (Tremblrel. 19, Created)  
 DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)  
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)  
 DE Hypothetical protein.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 CC NCBI\_TaxID=10090;  
 RX MEDLINE=96071149; PubMed=7475288;  
 RA Cao J., Vescio R.A., Rettig M.B., Hong C.H., Kim A., Lee J.C.,  
 RA Lichtenstein A.K., Berenson J.R.,  
 RT "A CD10-positive subset of malignant cells is identified in multiple  
 RT leukemia using PCR with patient-specific immunoglobulin gene primers."  
 RL Leukemia 9:1948-1953(1995).  
 DR EMBL; S80860; AAD14339.1; -.  
 DR HSSP; P01772; 2FB4.  
 DR GO; GO:0005887; C:integral to plasma membrane; NAS.  
 DR GO; GO:0016489; F:immunoglobulin receptor activity; NAS.  
 DR GO; GO:0016066; P:cellular defense response (sensu Vertebrata); NAS.  
 DR InterPro; IPRO07110; IG-like.  
 DR Pfam; PF00047; Ig; 1.  
 DR SMART; SM00406; IgV; 1.  
 DR PROSITE; PS50835; IG-LIKE; 4.  
 DR PROSITE; PS50835; IG-LIKE; 2.  
 KW Hypothetical protein.  
 SQ SEQUENCE 479 AA; 51603 MW; ECE2D0877748564F CRC64;  
 Query Match 50.8%; Score 348; DB 11; Length 479;  
 Best Local Similarity 58.8%; Pred. No. 6e-26;  
 Matches 73; Conservative 17; Mismatches 26; Indels 8; Gaps 3;  
 QY 1 QVQLQESGGGLVQAGSRLSCAASGSIQSIHTMGWRQTGKERDVATITQDGSSTN-Y 59  
 DB 1 QVHVLVESGGGVQVQPKSLRLSCASGFTFTSTYGMWVRQAPGKGLDWALISYDGSSTQY 60  
 QY 60 ADSVKGRTISRDNLTNTVYLQNMNDLKPEDTAVYYCNADVPRPYTSRY---LEWVGQGT 119  
 DB 61 ADSVKGRTISRDNKNTLYLQMSLRVEDIATVYCAADGVPFVSQVYVAGIYWGQGT 132  
 QY 120 SSEP 123  
 DB 133 SSEP 136  
 RESULT 14  
 Q8WU38 PRELIMINARY; PRT; 573 AA.  
 ID Q8WU38;  
 AC Q8WU38;  
 DT 01-MAR-2002 (Tremblrel. 20, Created)  
 DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)

DT	01-OCF-2003 (T-EMBLrel. 25, Last annotation update)
DE	Hypothetical protein.
OS	Homo sapiens (human).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX	NCBI_Taxid=9606;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	TISSUE=Tonsil;
RA	Strausberg R.;
RL	Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR	EMBL; BC021276; AAH21276.1; -
DR	PIR; S21205; S21205.
DR	InterPro; IPR007110; Ig_1ike.
DR	InterPro; IPR003006; IG_MHC.
DR	InterPro; IPR003596; Ig_v.
DR	Pfam; PF00047; Ig_4.
DR	SMART; SM00406; IgY_1.
DR	PROSITE; PSS0835; IG_LIKE; 4.
DR	PROSITE; PS00230; IG_MHC; 2.
KW	Hypothetical protein.
SQ	SEQUENCE 573 AA; 62967 MW; FD072344033AC530 CRC64;
Query March	50.8%; Score 348; DB 4; Length 573;
Blast Local Similarity	56.5%; Pred No. 7,66-26;
Matches 78; Conservative 13; Mismatches 35; Indels 12; Gaps 4	
Dc	1 QVQLAQSGLGVAGQSLRLSCAASGSGISHTHWGYRGQTGERDYVAITQ-DGGSTNY 59
Dc	20 EVOLVSSGGVLVPGKSHRLSCAASGFPPDDYAAMHWRAQPKGLEVSSGISWNSSIGY 79
Dc	60 ADVVKRFPTSPNTLNITYLLCNMLKEPDPAVVYCAVDREPTSY-----LEWCG 113
Dc	80 ADVSKGFPTISRNNANSLYLQNSLRADFDALTYC--AKHGSGSIGIYYGNWDWGQ 135
Dc	114 GLTVVSSEP-KTPKEQP 130
Dc	136 GTVTVSSAPTKAPDVEP 153
RESULT 15	
ID	Q9UL91 PRELIMINARY; PRT; 118 AA.
AC	Q9UL91;
DT	01-MAY-2000 (T-EMBLrel. 13, Created)
DT	01-MAY-2000 (T-EMBLrel. 13, Last sequence update)
DT	01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE	Myosin-reactive immunoglobulin heavy chain variable region (Fragment).
OS	Homo sapiens (human).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX	NCBI_Taxid=9606;
RN	[1]
RP	SEQUENCE FROM N.A.
RL	MEDLINE=98277139; PubMed=9614934; Kall's N.N., Berney S.M., Young D.C.;
RA	Young D.C.;
RT	"Myosin-reactive autoantibodies in rheumatic carditis and normal fetuss";
RL	Clin. Immunol. Immunopathol. 87:184-192(1998).
DR	EMBL; AF035023; AAD56259.1; -
DR	PIR; S21205; S21205.
DR	HSSP; P01772; 2PB4.
DR	InterPro; IPR007110; Ig_1like.
DR	InterPro; IPR003596; Ig_v.
DR	Pfam; PF00047; Ig_1.
DR	SMART; SM00406; IgY_1.
DR	PROSITE; PSS0835; IG_LIKE; 1.
FT	NON_TER 1
FT	NON_TER 118
SEQ	SEQUENCE 118 AA; 12843 MW; D0633949F2AC149D CRC64;

Query Match Similarity	50.4%	Score 345.5	DB 4	Length 118
Best Local Similarity	61.5%	Pred. No. 1,76-26		
Matches	75	Conservative 12	Mismatches 28	Indels 7
			Gaps 3	

  

Qy	1	QVQLQESGGGLVQAGSRLRSCAASGISLHTGWYRQPKGERDVAITDGGST	NY	59
	:	.....	:	:
Db	1	EVLVLSGGGLVQPGGSLRLSCAASGFFPSYSMMWRQAPRGKLEWYSYISITITYY		60
	:	.....	:	:
Qy	60	ADSVKRRFTISDNTLNTVYLQANDLKPEPTAYVYC-NADVRYKRSRLIEWGGT	LT	118
	:	.....	:	:
Db	61	ADSVKRRFTISDNTAKNSLYLQNNLSRAEDPAAVYVCARGD-----SSAEDFIDWGQIMVT		115
	:	.....	:	:
Qy	119	VS	120	
	:	:	:	
Db	116	VS	117	

Search completed: October 5, 2004, 08:13:44  
Job time : 69.817 secs



OM protein - protein search, using sw model

Run on: October 5, 2004, 07:04:10 ; Search time 98.9043 Seconds

371.381 Million cell updates/sec

Title: US-09-805-290A-19

Sequence: 1 QVQLQESGGGLVQAGGSLRL...WGQGTLLVTSSEPKTKPKPQ 130

Scoring table: BLOSUM62

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_29Jan04:\*

```

1: _A_Geneseq_29Jan04:*
2: 1: geneseq11980s:*
3: 2: geneseq11990s:*
4: 3: geneseq2000s:*
5: 4: geneseq2001s:*
6: 5: geneseq2002s:*
7: 6: geneseq2003as:*
8: 7: geneseq2003bs:*
9: 8: geneseq2004s:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	685	100.0	130	4	AAE10552	AAe10552 HPL inhib
2	659	96.2	130	4	AAE10554	AAe10554 HPL inhib
3	641	93.6	130	4	AAE10559	AAe10559 HPL inhib
4	532.5	77.7	131	4	AAE10555	AAe10555 HPL inhib
5	522	76.2	130	4	AAE10556	AAe10556 HPL inhib
6	513	74.9	130	4	AAE10557	AAe10557 HPL inhib
7	500.5	73.1	129	4	AAE10553	AAe10553 HPL inhib
8	497	72.6	152	4	AAE05283	AAe05283 Anti-porc
9	490	71.5	124	4	AAE10563	AAe10563 HPL inhib
10	485.5	70.9	131	5	ABG30620	ABg30620 Immunoglo
11	485	70.8	124	4	AAE10566	AAe10566 HPL inhib
12	479.5	70.0	129	4	AAE10558	AAe10558 HPL inhib
13	472.5	69.0	129	4	AAE10561	AAe10561 HPL inhib
14	470.5	68.7	149	4	AAE05288	AAe05288 Anti-porc
15	462	68.5	130	4	AAE10567	AAe10567 HPL inhib
16	468.5	68.4	129	4	AAE10564	AAe10564 HPL inhib
17	463	67.6	128	4	AAE10565	AAe10565 HPL inhib
18	455	66.4	190	7	ABR62883	ABr62883 Liama ant
19	54.5	66.4	129	4	AAE10551	AAe10551 HPL inhib
20	453.5	66.2	118	2	AAAY3819	AAy3819 Liama ant
21	453	66.1	190	7	ABR62882	ABr62882 Liama ant
22	452.5	66.1	377	4	AAE05286	AAe05286 Liama ant
23	452.5	66.1	383	4	AAE05287	AAe05287 Liama HCV
24	451	65.8	190	7	ABR62881	ABr62881 Liama ant
25	449	65.5	190	7	ABR62879	ABr62879 Liama ant

26	449	65.5	198	7	ABR62878	Llama ant
27	448	65.4	190	7	ABR62880	Abro62880 Llama ant
28	444.5	64.9	211	2	AAy41172	AAy41172 Llama Vhh
29	443	64.7	134	5	ABG30618	ABG30618 Immunoglob
30	442.5	64.6	204	2	AAy41167	AAy41167 Llama Vhh
31	439	64.1	119	4	ABE67778	ABE67778 Llama Vhh
32	439	64.1	119	4	ABE67777	Amino aci
33	43.5	63.4	153	4	AAE05282	Ant1-PolA
34	434.5	63.4	225	2	AAy41165	Llama Vhh
35	431.5	63.1	205	2	AAy41173	Llama Vhh
36	43.5	62.9	194	7	ABR62885	ABR62885 Llama ant
37	429	62.6	119	4	ABG67779	Amino aci
38	428.5	62.6	133	5	ABG30972	Immunoglob
39	427.5	62.4	124	2	AAy28615	Llama ant
40	425	62.0	111	2	AAy39816	Llama ant
41	425	62.0	124	4	AAE10562	HGL inhib
42	424	61.9	125	2	AAy39817	Llama ant
43	423.5	61.7	127	5	ABG30617	Mouse cro
44	422.5	61.7	193	7	ABR62884	AbG62884 Llama ant
45	422	61.6	130	4	ABE10560	HGL inhib

## ALIGNMENTS

## RESULT 1

ID	AAE10552	standard; peptide; 130 AA.
1		
2		
3		
4		
5		
6		
7		
8		
9		
10		
11		
12		
13		
14		
15		
16		
17		
18		
19		
20		
21		
22		
23		
24		
25		
26		
27		
28		
29		
30		
31		
32		
33		
34		
35		
36		
37		
38		
39		
40		
41		
42		
43		
44		
45		
46		
47		
48		
49		
50		
51		
52		
53		
54		
55		
56		
57		
58		
59		
60		
61		
62		
63		
64		
65		
66		
67		
68		
69		
70		
71		
72		
73		
74		
75		
76		
77		
78		
79		
80		
81		
82		
83		
84		
85		
86		
87		
88		
89		
90		
91		
92		
93		
94		
95		
96		
97		
98		
99		
100		

AC AA610552;

DT 10-DEC-2001 (first entry)

HPL inhibiting VHH fragment, HPL #12 from llama species

KW Llama antobody; camelid; anorectic; heavy chain variable domain; VHH; human diet restriction inhibitor; human anorectic lipase; HPI

KW Food; human gastric lipase; HGL; cosmetic control; body weight.

Lama sp.

EH	Key	Location/Qualifiers
FT		31 35

```

ET      /label= CDK1
ET      /note= "Complementarity determining region 1"

```

ET	Region	50 : .64	
ET	/label= CDR2		

```

FI /note= "complementarily determining region 2
FT 98, .110
Region

```

```

E1      /label= CDR3
FT      /note= "Complementarity determining region 3"

```

AA  
PN  
EP1134231-A1

PD 19-SEP-2001

20-FEB-2001; 2001EP-00200703.

PR 14-MAR-2000; 2000EP-00200930.

PA (UNIT ) UNILEVER NV.

XX

XX

XX

PT useful for cosmetic

PT of light chains.

PS Example 2; Page 9; 3

Example 2; Page 9; 37pp; English.

XX The patent discloses antibodies or their fragments comprising a heavy  
 CC chain variable domain (VH) derived from an immunoglobulin naturally  
 CC devoid of light chains specific for inhibiting human dietary enzymes. The  
 CC antibodies of the invention are useful for the preparation of medicaments  
 CC or food for inhibiting the activity of one or more human dietary enzymes  
 CC especially human pancreatic lipase (HPL) or human gastric lipase (HGL)  
 CC which are useful for the cosmetic control of body weight of human beings.  
 CC The present peptide sequence is HPL inhibiting VH fragment, HPL #12 from  
 CC llama (camelid) species  
 XX  
 SQ Sequence 130 AA;  
 Query Match 100.0%; Score 685; DB 4; Length 130;  
 Best Local Similarity 100.0%; Pred. No. 8.1e-52;  
 Matches 130; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 QVQLQESGGGLVQAGSLRLSCAASGSIHTMGYRQTPGKERDVAVTIIDGGSTNYA 60  
 DB 1 QVQLQESGGGLVQAGSLRLSCAASGSIHTMGYRQTPGKERDVAVTIIDGGSTNYA 60  
 QY 61 DSVKGRFTISRDTLTNTVYLQWMDLKPEDTAVYYCNADVPRPRTSRYLEWVGQGLTVTVS 120  
 DB 61 DSVKGRFTISRDTLTNTVYLQWMDLKPEDTAVYYCNADVPRPRTSRYLEWVGQGLTVTVS 120  
 QY 121 SEPTKPKPP 130  
 DB 121 SEPTKPKPP 130  
 RESULT 2  
 AAE10554  
 ID AAE10554 standard; peptide; 130 AA.  
 XX  
 AC AAE10554;  
 XX  
 DT 10-DEC-2001 (first entry)  
 XX  
 DE HPL inhibiting VH fragment, HPL #14 from llama species.  
 XX  
 KM llama antibody; camelid; anorectic; heavy chain variable domain; VH;  
 KW human dietary enzyme inhibitor; medicament; human pancreatic lipase; HPL;  
 KM food; human gastric lipase; HGL; cosmetic control; body weight.  
 XX  
 OS Lama sp.  
 XX  
 FH Key Location/Qualifiers  
 FT Region /label= CDR1  
 FT /note= "Complementarity determining region 1"  
 FT 50..64  
 FT /label= CDR2  
 FT /note= "Complementarity determining region 2"  
 FT 98..110  
 FT /label= CDR3  
 FT /note= "Complementarity determining region 3"  
 FT  
 XX  
 XX EP1134231-A1.  
 XX  
 PD 19-SEP-2001.  
 XX  
 PF 20-FEB-2001; 2001EP-00200703.  
 XX  
 PR 14-MAR-2000; 2000EP-00200930.  
 XX  
 PA (UNIL ) UNILEVER NV.  
 PA (UNIL ) UNILEVER PLC.  
 XX  
 PI Bezemer S, Van De Burg M, De Haard JWM, Tareilus E;  
 DR WPI; 2001-572718/65.  
 XX  
 PT New antibody or its fragments for inhibiting human dietary enzymes,

PT useful for cosmetic control of body weight of human beings, comprises  
 PT heavy chain variable domain derived from immunoglobulin naturally devoid  
 PT of light chains.  
 XX  
 PS Example 2; Page 9; 37pp; English.  
 XX  
 CC The patent discloses antibodies or their fragments comprising a heavy  
 CC chain variable domain (VH) derived from an immunoglobulin naturally  
 CC devoid of light chains specific for inhibiting human dietary enzymes. The  
 CC antibodies of the invention are useful for the preparation of medicaments  
 CC or food for inhibiting the activity of one or more human dietary enzymes  
 CC especially human pancreatic lipase (HPL) or human gastric lipase (HGL)  
 CC which are useful for the cosmetic control of body weight of human beings.  
 CC The present peptide sequence is HPL inhibiting VH fragment, HPL #14 from  
 CC llama (camelid) species  
 XX  
 SQ Sequence 130 AA;  
 Query Match 96.2%; Score 659; DB 4; Length 130;  
 Best Local Similarity 96.2%; Pred. No. 1.5e-49;  
 Matches 125; Conservative 2; Mismatches 3; Indels 0; Gaps 0;  
 QY 1 QVQLQESGGGLVQAGSLRLSCAASGSIHTMGYRQTPGKERDVAVTIIDGGSTNYA 60  
 DB 1 QVQLQESGGGLVQAGSLRLSCAASGSIHTMGYRQTPGKERDVAVTIIDGGSTNYA 60  
 QY 61 DSVKGRFTISRDTLTNTVYLQWMDLKPEDTAVYYCNADVPRPRTSRYLEWVGQGLTVTVS 120  
 DB 61 DSVKGRFTISRDTLTNTVYLQWMDLKPEDTAVYYCNADVPRPRTSRYLEWVGQGLTVTVS 120  
 QY 121 SEPTKPKPP 130  
 DB 121 SEPTKPKPP 130  
 RESULT 3  
 AAE10559  
 ID AAE10559 standard; peptide; 130 AA.  
 XX  
 AC AAE10559;  
 XX  
 DT 10-DEC-2001 (first entry)  
 XX  
 DE HPL inhibiting VH fragment, HPL #30 from llama species.  
 XX  
 KM llama antibody; camelid; anorectic; heavy chain variable domain; VH;  
 KW human dietary enzyme inhibitor; medicament; human pancreatic lipase; HPL;  
 KM food; human gastric lipase; HGL; cosmetic control; body weight.  
 XX  
 OS Lama sp.  
 XX  
 FH Key Location/Qualifiers  
 FT Region /label= CDR1  
 FT /note= "Complementarity determining region 1"  
 FT 49..64  
 FT /label= CDR2  
 FT /note= "Complementarity determining region 2"  
 FT 98..110  
 FT /label= CDR3  
 FT /note= "Complementarity determining region 3"  
 FT  
 XX  
 XX EP1134231-A1.  
 XX  
 PD 19-SEP-2001.  
 XX  
 PF 20-FEB-2001; 2001EP-00200703.  
 XX  
 PR 14-MAR-2000; 2000EP-00200930.  
 XX  
 PA (UNIL ) UNILEVER NV.  
 PA (UNIL ) UNILEVER PLC.  
 XX  
 PT



```

XX 19-SEP-2001.
PD 20-FEB-2001; 2001EP-00200703.
XX PF
XX 14-MAR-2000; 2000EP-00200930.
XX PR
XX (UNIL ) UNILEVER NV.
XX PA (UNIL ) UNILEVER PLC.
XX PI
XX Bezemer S, Van De Burg M, De Haard JWM, Tareilus E;
XX WPI; 2001-572718/65.
XX DR
XX New antibody or its fragments for inhibiting human dietary enzymes,
PT useful for cosmetic control of body weight of human beings, comprises
PT heavy chain variable domain derived from immunoglobulin naturally devoid
PT of light chains.
XX PS
XX Example 2; Page 10; 37pp; English.
XX CC The patent discloses antibodies or their fragments comprising a heavy
CC chain variable domain (VHH) derived from an immunoglobulin naturally
CC devoid of light chains specific for inhibiting human dietary enzymes. The
CC antibodies of the invention are useful for the preparation of medicaments
CC or food for inhibiting the activity of one or more human dietary enzymes
CC especially human pancreatic lipase (HPL) or human gastric lipase (HGL)
CC which are useful for the cosmetic control of body weight of human beings.
CC The present peptide sequence is HPL inhibiting VHH fragment, HPL #18 from
CC llama (camelid) species
XX SQ
XX Sequence 130 AA;
XX Query Match 76.2%; Score 522; DB 4; Length 130;
XX Best Local Similarity 77.7%; Pred. No. 1.1e-37;
XX Matches 101; Conservative 9; Mismatches 20; Indels 0; Gaps 0;
XX QY 1 QVQLQESGGGLVQAGGSLRISCAASGSIHITMGWYRQTPGKERDVVATIDGGSTNYA 60
XX DB 1 QVQLQDSGGGLVQAGGSLRISCAASGTDIDYTMAMWRQAPGKERELVASATESSPNYA 60
XX QY 61 DSVKGRFTISRNDTLNLTNYIQLMNDLKPEDTAVYYCAADVPRTSRYLEVMGQGLTVTVS 120
XX DB 61 DSVKGRFTISRNDGKLTLYLQNMNLSKPEDTAVYYCNALIRKFTSEYNHWGQGIQVTVS 120
XX QY 121 SEPTKPKPP 130
XX DB 121 SEPTKPKPP 130
XX RESULT 6
XX AAE10557
XX ID AAE10557 standard; peptide; 130 AA.
XX AC AAE10557;
XX XX
XX DT 10-DEC-2001 (first entry)
XX DE HPL inhibiting VHH fragment, HPL #19 from llama species.
XX XX
XX llama antibody; camelid; anorectic; heavy chain variable domain; VHH;
XX human dietary enzyme inhibitor; medicament; human pancreatic lipase; HPL;
XX food; human gastric lipase; HGL; cosmetic control; body weight.
XX OS
XX Lama sp.
XX Key Location/Qualifiers
XX FH 31..35
XX FT /label= CDR1
XX FT /note= "Complementarity determining region 1"
XX FT 50..64
XX FT /label= CDR2
XX FT /note= "Complementarity determining region 2"
XX FT

```

```

FT Region 98..110
FT /label= CDR3
FT /note= "Complementarity determining region 3"
XX EP1134231-A1.
XX EN
XX 19-SEP-2001.
XX PD
XX 20-FEB-2001; 2001EP-00200703.
XX PF
XX 14-MAR-2000; 2000EP-00200930.
XX PR
XX (UNIL ) UNILEVER NV.
XX PA (UNIL ) UNILEVER PLC.
XX PI
XX Bezemer S, Van De Burg M, De Haard JWM, Tareilus E;
XX WPI; 2001-572718/65.
XX DR
XX New antibody or its fragments for inhibiting human dietary enzymes,
PT useful for cosmetic control of body weight of human beings, comprises
PT heavy chain variable domain derived from immunoglobulin naturally devoid
PT of light chains.
XX PS
XX Example 2; Page 10; 37pp; English.
XX CC The patent discloses antibodies or their fragments comprising a heavy
XX chain variable domain (VHH) derived from an immunoglobulin naturally
XX devoid of light chains specific for inhibiting human dietary enzymes. The
XX antibodies of the invention are useful for the preparation of medicaments
XX or food for inhibiting the activity of one or more human dietary enzymes
XX especially human pancreatic lipase (HPL) or human gastric lipase (HGL)
XX which are useful for the cosmetic control of body weight of human beings.
XX The present peptide sequence is HPL inhibiting VHH fragment, HPL #19 from
XX llama (camelid) species
XX SQ
XX Sequence 130 AA;
XX Query Match 74.9%; Score 513; DB 4; Length 130;
XX Best Local Similarity 75.4%; Pred. No. 6.9e-37;
XX Matches 98; Conservative 9; Mismatches 23; Indels 0; Gaps 0;
XX QY 1 QVQLQESGGGLVQAGGSLRISCAASGSIHITMGWYRQTPGKERDVVATIDGGSTNYA 60
XX DB 1 QVQLQDSGGGLVQAGGSLRISCAASGPIGDIYLMGWRQAPGKEREMVASITATGPNT 60
XX QY 61 DSVKGRFTISRNDTLNLTNYIQLMNDLKPEDTAVYYCAADVPRTSRYLEVMGQGLTVTVS 120
XX DB 61 DSVKGRFTISRNDKNTLEYLQNMNLSKPEDTAVYYCAOVAVRPSSDYTVYWGQGIQVTVS 120
XX QY 121 SEPTKPKPP 130
XX DB 121 SEPTKPKPP 130
XX RESULT 7
XX AAE10553
XX ID AAE10553 standard; peptide; 129 AA.
XX AC AAE10553;
XX XX
XX DT 10-DEC-2001 (first entry)
XX DE HPL inhibiting VHH fragment, HPL #13 from llama species.
XX XX
XX llama antibody; camelid; anorectic; heavy chain variable domain; VHH;
XX human dietary enzyme inhibitor; medicament; human pancreatic lipase; HPL;
XX food; human gastric lipase; HGL; cosmetic control; body weight.
XX OS
XX Lama sp.
XX Key Location/Qualifiers
XX FH 31..35
XX FT

```

```

FT FT /label= CDR1
FT FT /note= "Complementarity determining region 1"
FT FT /label= CDR2
FT FT /label= CDR2
FT FT /note= "Complementarity determining region 2"
FT FT /label= CDR3
FT FT /label= CDR3
FT FT /note= "Complementarity determining region 3"
PN EPI134231-A1.
XX PD 19-SEP-2001.
XX PF 20-FEB-2001; 2001EP-00200703.
XX PR 14-MAR-2000; 2000EP-00200930.
XX PA (UNIL ) UNILEVER NV.
XX PA (UNIL ) UNILEVER PLC.
XX PI Bezemer S, Van De Burg M, De Haard JW, Tareilus B;
XX DR WPI; 2001-572718/65.
XX PT New antibody or its fragments for inhibiting human dietary enzymes,
XX PT useful for cosmetic control of body weight of human beings, comprises
XX PT heavy chain variable domain derived from immunoglobulin naturally devoid
XX PT of light chains.
XX PS Example 2; Page 9; 37pp; English.
XX CC The patent discloses antibodies or their fragments comprising a heavy
XX CC chain variable domain (VH) derived from an immunoglobulin naturally
XX CC devoid of light chains specific for inhibiting human dietary enzymes. The
XX CC antibodies of the invention are useful for the preparation of medicaments
XX CC or food for inhibiting the activity of one or more human dietary enzymes
XX CC especially human pancreatic lipase (HPL) or human gastric lipase (HGL)
XX CC which are useful for the cosmetic control of body weight of human beings.
XX CC The present peptide sequence is HPL inhibiting VH fragment, HPL #13 from
XX CC llama (camelid) species
XX CC
XX SQ Sequence 129 AA;
Query Match 73.1%; Score 500.5; DB 4; Length 129;
Best Local Similarity 76.9%; Pred. No. 8.3e-36;
Matches 100; Conservative 8; Mismatches 21; Indels 1; Gaps 1;
QY 1 QVQLQESGGGLVQAGGSLRLSCAASGSIIGTWMGVRQPKGERDVVATIQDGGSTNYA 60
DB 1 QVQLQESGGGLVQAGGSLRLSCAASGSIIGTWMGVRQPKGERDVVATIQDGGSTNYA 60
QY 61 DSVKRRFTISRNTLTNTLYLQWMDLKPEDTAVYYCAAVRPRRSRYLEWVGGLTVVS 120
DB 61 DSAKRFTISKNNANTVYLQWNSLKPBDTAVYYCNALITRWKDS -VNDYGGGQGTQVTVS 119
QY 121 SEPKTPKPOP 130
DB 120 SEPKTPKPOP 129
XX
XX RESULT 8
XX ID AAE05283 standard; protein; 152 AA.
XX AC AAE05283;
XX DT 18-SEP-2001 (first entry)
XX XX Anti-potato SBEII (Clone46) VH region attached with myc and his6 tag.
XX DE Potato; heavy chain immunoglobulin; pathogen resistance;
XX XX metabolism modulator; passive immunisation; heavy chain variable domain;
XX KM VH; anti-potato SBEII; starch branching enzyme; SBE A.

```

```

XX XX Solanum tuberosum.
XX OS Unidentified.
XX OS Chimeric.
XX XX EPI118669-A2.
XX XX 25-JUL-2001.
XX PF 08-DEC-2000; 2000EP-00310997.
XX PR 17-DEC-1999; 99EP-00310188.
XX PA (UNIL ) UNILEVER PLC.
XX PA (UNIL ) UNILEVER NV.
XX XX Franken LGJ, Van Der Logt CPE, Jobling SA, Teh Y;
XX DR WPI; 2001-427157/46.
XX DR N-PSDB; AAD10054.
XX PT Modifying a plant to produce an antibody useful for increasing pathogen
XX PT resistance or to modulate metabolism comprises introducing a DNA sequence
XX PT encoding a heavy chain immunoglobulin linked to a peptide that targets a
XX PT cellular compartment.
XX PS Example 1; Fig 12; 81pp; English.
XX CC The present invention relates to a method for modifying a plant to
XX CC produce an antibody or an active fragment or derivative, or a protein
XX CC functional equivalent, in a cellular compartment comprising introducing a
XX CC DNA sequence encoding a heavy chain immunoglobulin, where the DNA is
XX CC linked to promoters and provided with an additional sequence encoding a
XX CC peptide capable of targeting heavy chain immunoglobulin to a cellular
XX CC compartment. The method is used for producing a heavy chain
XX CC immunoglobulin or an active fragment or derivative, or a protein that is
XX CC functionally equivalent for increasing the pathogen resistance in a plant
XX CC or to modulate metabolism in a plant. Under some circumstances it may be
XX CC desirable to retain the antibody product with the plant rather than
XX CC extracting and isolating the product. In particular, edible selected
XX CC antigens may be used in a method of passively immunising an animal,
XX CC preferably human, against the antigen, e.g., pathogenic organisms. The
XX CC present sequence is an anti-potato SBEII antibody (denoted Clone46) heavy
XX CC chain variable domain (VH) attached to peptide linkers, myc and his6 tag.
XX CC The potato SBEII is a starch branching enzyme also designated SBE A
XX CC
XX SQ Sequence 152 AA;
Query Match 72.6%; Score 497; DB 4; Length 152;
Best Local Similarity 75.4%; Pred. No. 2e-35;
Matches 101; Conservative 5; Mismatches 16; Indels 12; Gaps 2;
QY 1 QVQLQESGGGLVQAGGSLRLSCAASGSIIGTWMGVRQPKGERDVVATIQDGGSTNYA 60
DB 1 QVQLQESGGGLVQAGGSLRLSCAASGSIIGTWMGVRQPKGERDVVATIQDGGSTNYA 60
QY 61 DSVKRRFTISRNTLTNTLYLQWMDLKPEDTAVYYCAVAD----VRPRTRSYLEWVGGLTV 116
DB 61 DSVKRRFTISRNTLTNTLYLQWMDLKPEDTAVYYCAVAD----VRPRTRSYLEWVGGLTV 116
QY 117 VTVSSEPKTPKPOP 130
DB 113 VTVSSEPKTPKPOP 126
XX
XX RESULT 9
XX ID AAE10563 standard; peptide; 124 AA.
XX AC AAE10563;
XX DT 10-DEC-2001 (first entry)
XX XX

```



Db 1 QVQLQESGGGLVQPGGSLRLSCAASKSIFGFAVGMHROAPGKQELVARITYDSGNTYA 60  
QY 61 DSVKGRFTISRDNLTNTVYLQMNDLKPEDTAVYYCNADVPRYTSRYL-EVWGQGLTVTV 119  
Db 61 DSVKGRFTISRDNLTNTVYLQMNLSLKPEDTGYVYCNAEIVRATITGFTIDMGQGLTVTV 120  
QY 120 SESEPTPKRQP 130  
Db 121 SESEPTPKRQP 131

## RESULT 11

AAE10566  
ID AAE10566 standard; peptide; 124 AA.

AAE10566;

10-DEC-2001 (first entry)

HGL inhibiting VHH fragment, HGL #15 from llama species.

Llama antibody; camelid; anorectic; heavy chain variable domain; VHH;  
human dietary enzyme inhibitor; medicament; human pancreatic lipase; HPL;  
food; human gastric lipase; HGL; cosmetic control; body weight.

Llama sp.

Location/Qualifiers

30..35  
/label=CDR1  
/note="Complementarity determining region 1"

50..64  
/label=CDR2  
/note="Complementarity determining region 2"

98..104  
/label=CDR3  
/note="Complementarity determining region 3"

Region

EP1134231-A1.

19-SEP-2001.

20-FEB-2001; 2001EP-00200703.

14-MAR-2000; 2000EP-00200930.

(UNIL ) UNILEVER NV.

(UNIL ) UNILEVER PLC.

Bezemer S, Van De Burg M, De Haard JWM, Tareilus E,

WPI; 2001-572718/65.

New antibody or its fragments for inhibiting human dietary enzymes,  
useful for cosmetic control of body weight of human beings, comprises  
heavy chain variable domain derived from immunoglobulin naturally devoid  
of light chains.

Example 4; Page 14; 37pp; English.

The patent discloses antibodies or their fragments comprising a heavy  
chain variable domain (VHH) derived from an immunoglobulin naturally  
devoid of light chains specific for inhibiting human dietary enzymes. The  
antibodies of the invention are useful for the preparation of medicaments  
or food for inhibiting the activity of one or more human dietary enzymes  
especially human pancreatic lipase (HPL) or human gastric lipase (HGL)  
CC which are useful for the cosmetic control of body weight of human beings.  
The present peptide sequence is HGL inhibiting VHH fragment, HGL #15 from  
llama (camelid) species

Sequence 124 AA;

Query Match 70.8%; Score 485; DB 4; Length 124;

Best Local Similarity 73.1%; Pred. No. 1.8e-34;  
Matches 95; Conservative 6; Mismatches 23; Indels 6; Gaps 1;

QY 1 QVQLQESGGGLVQAGSLRLSCAASGSGSIHMMGWRCRPGKERDVATIQGGSTNYA 60  
Db 1 QVQLQESGGGLVQAGSLRLSCAASGSGSIHMMGWRCRPGKERDVATIQGGSTNYA 60  
QY 61 DSVKGRFTISRDNLTNTVYLQMNDLKPEDTAVYYCNADVPRYTSRYL-EVWGQGLTVTVS 120  
Db 61 DSVKGRFTISRDNLTNTVYLQMNLSLKPEDTGYVYCNAEIVRATITGFTIDMGQGLTVTVS 114  
QY 121 SESEPTPKRQP 130  
Db 115 SESEPTPKRQP 124

## RESULT 12

AAE10558  
ID AAE10558 standard; peptide; 129 AA.

AAE10558;

10-DEC-2001 (first entry)

HPL inhibiting VHH fragment, HPL #22 from llama species.

Llama antibody; camelid; anorectic; heavy chain variable domain; VHH;  
human dietary enzyme inhibitor; medicament; human pancreatic lipase; HPL;  
food; human gastric lipase; HGL; cosmetic control; body weight.

Llama sp.

Location/Qualifiers

31..35  
/label=CDR1  
/note="Complementarity determining region 1"

50..64  
/label=CDR2  
/note="Complementarity determining region 2"

98..109  
/label=CDR3  
/note="Complementarity determining region 3"

Region

EP1134231-A1.

19-SEP-2001.

20-FEB-2001; 2001EP-00200703.

14-MAR-2000; 2000EP-00200930.

(UNIL ) UNILEVER NV.

(UNIL ) UNILEVER PLC.

Bezemer S, Van De Burg M, De Haard JWM, Tareilus E,

WPI; 2001-572718/65.

New antibody or its fragments for inhibiting human dietary enzymes,  
useful for cosmetic control of body weight of human beings, comprises  
heavy chain variable domain derived from immunoglobulin naturally devoid  
of light chains.

Example 2; Page 10; 37pp; English.

The patent discloses antibodies or their fragments comprising a heavy  
chain variable domain (VHH) derived from an immunoglobulin naturally  
devoid of light chains specific for inhibiting human dietary enzymes. The  
antibodies of the invention are useful for the preparation of medicaments  
or food for inhibiting the activity of one or more human dietary enzymes  
especially human pancreatic lipase (HPL) or human gastric lipase (HGL)  
CC which are useful for the cosmetic control of body weight of human beings.  
The present peptide sequence is HPL inhibiting VHH fragment, HPL #22 from

CC llama (camelid) species  
 XX  
 SQ Sequence 129 AA;  
 Query Match 70.0%; Score 479.5; DB 4; Length 129;  
 Best Local Similarity 73.8%; Pred. No. 5.5e-34;  
 Matches 96; Conservative 10; Mismatches 23; Indels 1; Gaps 1;  
 Db 1 QVQLQESGGGLVQAGGSLRLSCAASGSIQSIHTMGWYRQTPGKERDVAATIDGGSTNYA 60  
 1 QVQLQESGGGLVQAGGSLRLSCAASGSIQSIHTMGWYRQTPGKERDVAATIDGGSTNYA 60  
 QY 61 DSVKGRFTISRDNTLTNTVYLQWMDLKEPTAVYVCNADYRPTSRYLEWVGQGLTVYS 120  
 61 DSVKGRFTISRDNTLTNTVYLQWMDLKEPTAVYVCNADYRPTSRYLEWVGQGLTVYS 120  
 Db 61 DSVKGRFTISRDNTLTNTVYLQWMDLKEPTAVYVCNADYRPTSRYLEWVGQGLTVYS 119  
 61 DSVKGRFTISRDNTLTNTVYLQWMDLKEPTAVYVCNADYRPTSRYLEWVGQGLTVYS 119  
 QY 121 SEKTPKPOP 130  
 121 SEKTPKPOP 130  
 Db 120 SEKTPKPOP 129  
 120 SEKTPKPOP 129  
 RESULT 13  
 AAE10561  
 ID AAE10561 standard; peptide; 129 AA.  
 XX  
 AC AAE10561;  
 XX  
 DT 10-DEC-2001 (first entry)  
 XX  
 DE HGL inhibiting VHH fragment, HGL #4 from llama species.  
 XX  
 KM llama antibody; camelid; anorectic; heavy chain variable domain; VHH;  
 KM human dietary enzyme inhibitor; medicament; human pancreatic lipase; HPL;  
 KM food; human gastric lipase; HGL; cosmetic control; body weight.  
 OS  
 OS Lama sp.  
 FH Location/Qualifiers  
 FH Key 31..35  
 FT Region /label= CDR1  
 FT /note= "Complementarity determining region 1"  
 FT 50..64  
 FT /label= CDR2  
 FT /note= "Complementarity determining region 2"  
 FT 98..109  
 FT Region /label= CDR3  
 FT /note= "Complementarity determining region 3"  
 PN EPI134231-A1.  
 XX  
 XX 19-SEP-2001.  
 XX  
 XX 20-FEB-2001; 2001EP-00200703.  
 XX  
 XX 14-MAR-2000; 2000EP-00200930.  
 XX  
 XX (UNITL ) UNILEVER NV.  
 XX (UNITL ) UNILEVER PLC.  
 XX  
 XX Bezemer S, Van De Burg M, De Haard JWM, Tareilus E;  
 XX WPI; 2001-572718/65.  
 DR  
 XX  
 XX New antibody or its fragments for inhibiting human dietary enzymes,  
 PT useful for cosmetic control of body weight of human beings, comprises  
 PT heavy chain variable domain derived from immunoglobulin naturally devoid  
 PT of light chains.  
 XX  
 XX Example 4; Page 13; 37pp; English.  
 CC The patent discloses antibodies or their fragments comprising a heavy  
 CC chain variable domain (VHH) derived from an immunoglobulin naturally  
 CC devoid of light chains specific for inhibiting human dietary enzymes. The

CC antibodies of the invention are useful for the preparation of medicaments  
 CC or food for inhibiting the activity of one or more human dietary enzymes  
 CC especially human pancreatic lipase (HPL) or human gastric lipase (HGL)  
 CC which are useful for the cosmetic control of body weight of human beings.  
 CC The present peptide sequence is HGL inhibiting VHH fragment, HGL #4 from  
 CC llama (camelid) species  
 XX  
 SQ Sequence 129 AA;  
 Query Match 69.0%; Score 472.5; DB 4; Length 129;  
 Best Local Similarity 71.2%; Pred. No. 2.2e-33;  
 Matches 94; Conservative 10; Mismatches 23; Indels 5; Gaps 2;  
 Db 1 QVQLQESGGGLVQAGGSLRLSCAASGSIQSIHTMGWYRQTPGKERDVAATIDGGSTNYA 60  
 1 QVQLQESGGGLVQAGGSLRLSCAASGSIQSIHTMGWYRQTPGKERDVAATIDGGSTNYA 60  
 QY 61 DSVKGRFTISRDNTLTNTVYLQWMDLKEPTAVYVCNADYRPTSRYLEWVGQGLTVY 118  
 61 DSVKGRFTISRDNTLTNTVYLQWMDLKEPTAVYVCNADYRPTSRYLEWVGQGLTVY 118  
 Db 61 DSVKGRFTISRDNTLTNTVYLQWMDLKEPTAVYVCNADYRPTSRYLEWVGQGLTVY 117  
 61 DSVKGRFTISRDNTLTNTVYLQWMDLKEPTAVYVCNADYRPTSRYLEWVGQGLTVY 117  
 QY 119 VSSSEKTPKPOP 130  
 119 VSSSEKTPKPOP 130  
 Db 118 VSSSEKTPKPOP 129  
 118 VSSSEKTPKPOP 129  
 RESULT 14  
 AAE05288  
 ID AAE05288 standard; protein; 149 AA.  
 XX  
 AC AAE05288;  
 XX  
 DT 18-SEP-2001 (first entry)  
 XX  
 DE Anti-potato SBHII (Clone68) VH region attached with myc and his6 tag.  
 XX  
 KM Potato; heavy chain immunoglobulin; pathogen resistance;  
 KM metabolism modulator; passive immunisation; heavy chain variable domain;  
 KM VH; anti-potato SBHII; starch branching enzyme; SBE A.  
 OS  
 OS Solanum tuberosum.  
 OS Unidentified.  
 OS Chimeric.  
 PN EPI118669-A2.  
 XX  
 XX 25-JUN-2001.  
 XX  
 XX 08-DEC-2000; 2000EP-00310997.  
 XX  
 XX 17-DEC-1999; 99EP-00310188.  
 XX  
 XX (UNITL ) UNILEVER PLC.  
 XX (UNITL ) UNILEVER NV.  
 XX  
 XX Franken LGJ, Van Der Logt CPE, Jobling SA, Teh Y;  
 XX WPI; 2001.427157/46.  
 DR N-PSDB; AAD10059.  
 XX  
 XX Modifying a plant to produce an antibody useful for increasing pathogen  
 PT resistance or to modulate metabolism comprises introducing a DNA sequence  
 PT encoding a heavy chain immunoglobulin linked to a peptide that targets a  
 PT cellular compartment.  
 XX  
 XX Example 13; Fig 28; 81pp; English.  
 CC The present invention relates to a method for modifying a plant to  
 CC produce an antibody or an active fragment or derivative, or a protein  
 CC functional equivalent, in a cellular compartment comprises introducing a  
 CC DNA sequence encoding a heavy chain immunoglobulin, where the DNA is  
 CC linked to promoters and provided with an additional sequence encoding a  
 CC peptide capable of targeting heavy chain immunoglobulin to a cellular



CC compartment. The method is used for producing a heavy chain  
 CC immunoglobulin or an active fragment or derivative, or a protein that is  
 CC functionally equivalent for increasing the pathogen resistance in a plant  
 CC or to modulate metabolism in a plant. Under some circumstances it may be  
 CC desirable to retain the antibody product with the plant rather than  
 CC extracting and isolating the product. In particular, edible selected  
 CC antigens may be used in a method of passively immunising an animal.  
 CC preferably human, against the antigen, e.g., pathogenic organisms. The  
 CC present sequence is an anti-potato SBEII antibody (denoted clones6) heavy  
 CC chain variable domain (VH) attached to peptide linkers, myc and his6 tag.  
 CC The potato SBEII is a starch branching enzyme also designated SBE A

XX Sequence 149 AA;

Query Match 68.7%; Score 470.5; DB 4; Length 149;  
 Best Local Similarity 73.1%; Pred. No. 3.9e-33;  
 Matches 95; Conservative 6; Mismatches 20; Indels 9; Gaps 1;

QY 1 QVQLQESGGGLVQAGSIRLSCAASGSIHTMGWYRQTGKERDVAITIQDGSNTYA 60  
 DB 3 QVQLQESGGGLVQAGSIRLSCAASGSIHTMGWYRQTGKERDVAITIQDGSNTYA 62  
 QY 61 DSVKGRFTISRDNLTNTVYLQWMDLKPEDTAVYYCNADVPRPYRTSRYLEVWGQGLTVTS 120  
 DB 63 DSVKGRFTISRDNLTNTVYLQWMDLKPEDTAVYYCNADVPRPYRTSRYLEVWGQGLTVTS 113  
 QY 121 SEPKTPKPOP 130  
 DB 114 SEPKTPKPOP 123

RESULT 15

AAE10567 standard; peptide; 130 AA.

XX AAE10567;

DT 10-DEC-2001 (first entry)

DE HGL inhibiting VHH fragment, HGL #16 from llama species.

KW llama antibody; camelid; anorectic; heavy chain variable domain; VHH;  
 KW human dietary enzyme inhibitor; medicament; human pancreatic lipase; HPL;  
 KW food; human gastric lipase; HGL; cosmetic control; body weight.

XX Lama sp.

XX Key Location/Qualifiers

FT 31..35

FT /label= CDR1

FT /note= "Complementarity determining region 1"

FT 50..64

FT /label= CDR2

FT /note= "Complementarity determining region 2"

FT 98..110

FT /label= CDR3

FT /note= "Complementarity determining region 3"

XX EPI134231-A1.

XX 19-SEP-2001.

XX 20-FEB-2001; 2001EP-00200703.

XX 14-MAR-2000; 2000EP-00200930.

XX (UNIL ) UNILEVER NV.

XX (UNIL ) UNILEVER PLC.

XX Bezemer S, Van De Burg M, De Haard JWW, Tareilus E;

XX WPI; 2001-572718/65.

PT New antibody or its fragments for inhibiting human dietary enzymes,  
 PT useful for cosmetic control of body weight of human beings, comprises  
 PT heavy chain variable domain derived from immunoglobulin naturally devoid  
 PT of light chains.

XX Example 4; Page 14; 37pp; English.

PS The patent discloses antibodies or their fragments comprising a heavy  
 CC chain variable domain (VHH) derived from an immunoglobulin naturally  
 CC devoid of light chains specific for inhibiting human dietary enzymes. The  
 CC antibodies of the invention are useful for the preparation of medicaments  
 CC or food for inhibiting the activity of one or more human dietary enzymes  
 CC especially human pancreatic lipase (HPL) or human gastric lipase (HGL)  
 CC which are useful for the cosmetic control of body weight of human beings.  
 CC The present peptide sequence is HGL inhibiting VHH fragment, HGL #16 from  
 CC llama (camelid) species

XX Sequence 130 AA;

Query Match 68.5%; Score 469; DB 4; Length 130;  
 Best Local Similarity 73.8%; Pred. No. 4.5e-33;  
 Matches 96; Conservative 4; Mismatches 30; Indels 0; Gaps 0;

QY 1 QVQLQESGGGLVQAGSIRLSCAASGSIHTMGWYRQTGKERDVAITIQDGSNTYA 60  
 DB 1 QVQLQESGGGLVQAGSIRLSCAASGSIHTMGWYRQTGKERDVAITIQDGSNTYA 60  
 QY 61 DSVKGRFTISRDNLTNTVYLQWMDLKPEDTAVYYCNADVPRPYRTSRYLEVWGQGLTVTS 120  
 DB 61 DSVKGRFTISRDNLTNTVYLQWMDLKPEDTAVYYCNADVPRPYRTSRYLEVWGQGLTVTS 120  
 QY 121 SEPKTPKPOP 130  
 DB 121 SEPKTPKPOP 130

Search completed: October 5, 2004, 08:00:10  
 Job time : 99.9043 secs

This Page Blank (uspio)

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd

OM protein - protein search, using sw model

```
Run on:      October 5, 2004, 07:36:16 ; Search time 20.9028 Seconds
              (without alignments)
              593.639 Million cell updates/sec
```

Title: US-09-805-290A-20

Sequence: 1 QVQLQESGGGLVQAGSRL...WGQGTVTVSSBPKTKPKPQ 129

Scoring table: BLOSUM62

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%

Database : PIR\_78:\*

```
1: p1r1: *
2: p1r2: *
3: p1r3: *
4: p1r4: *
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	36.5	58.4	120	2	S5556	1g heavy chain V I
2	39.2.5	57.8	121	2	H36005	1g heavy chain V I
3	39.1.5	57.7	123	2	S31114	1g heavy chain - I
4	39.9.5	57.4	120	2	S55517	1g heavy chain V I
5	38.9.5	57.4	121	2	S26708	1g heavy chain V I
6	38.9.5	57.4	123	2	A36006	1g heavy chain V I
7	38.9	57.3	116	2	S31110	1g heavy chain - I
8	38.8.5	57.2	119	2	C36005	1g heavy chain - I
9	38.8	57.1	120	2	S48798	1g heavy chain V I
10	38.7.5	57.1	121	2	I55673	1g heavy chain V I
11	38.5.5	56.8	120	2	S55559	1g heavy chain - I
12	38.4.5	56.6	117	2	S31109	1g heavy chain - I
13	38.4.5	56.6	119	2	S31108	1g heavy chain - I
14	38.4	56.6	114	2	S31120	1g heavy chain - I
15	38.2.5	56.3	119	2	D36005	1g heavy chain V I
16	38.2.5	56.3	119	2	S31107	1g heavy chain - I
17	38.2.5	56.3	160	2	S05271	1g heavy chain - I
18	38.1.5	56.2	139	2	S31678	1g heavy chain pre
19	38.1.5	56.2	140	2	S31666	1g heavy chain V I
20	38.0.5	56.0	135	2	S31598	1g heavy chain V I
21	38.0	56.0	128	2	S26786	1g heavy chain V I
22	37.9.5	55.9	121	2	S31106	1g heavy chain - I
23	37.8	55.7	114	2	S31105	1g heavy chain (su
24	37.7.5	55.6	134	2	S31699	1g heavy chain V I
25	37.5.5	55.3	123	2	S26794	1g heavy chain V I
26	37.5.5	55.2	140	2	S31568	1g heavy chain V I
27	37.4.5	55.2	120	2	S55558	1g heavy chain V I
28	37.3.5	55.0	125	2	S30511	1g heavy chain V I
29	37.3.5	55.0	138	2	S31666	1g heavy chain V I

30	373	54.9	143	2	S2362	Ig heavy chain V r
31	372.5	54.9	121	2	G35005	Ig heavy chain V r
32	372	54.8	119	2	F27888	Ig heavy chain V r
33	372	54.8	122	2	S20782	Ig heavy chain V r
34	370.5	54.6	122	2	E27888	Ig heavy chain V r
35	370.5	54.6	152	2	B25471	Ig heavy chain pre
36	369.5	54.4	117	2	S78486	Ig heavy chain V r
37	369	54.3	128	2	S26790	Ig heavy chain V r
38	368	54.2	114	2	S45391	Ig heavy chain V r
39	367.5	54.1	138	2	A30561	Ig heavy chain pre
40	366.5	54.0	127	2	S38489	Ig heavy chain - h
41	366.5	54.0	140	2	S22657	Ig heavy chain pre
42	366	53.9	122	2	S20772	Ig heavy chain V r
43	366	53.9	137	2	S78054	Ig heavy chain pre
44	365.5	53.8	134	2	S31679	Ig heavy chain V r
45	365.5	53.8	139	2	I37781	Ig variable region

## ALIGNMENTS

## RESULT 1

Ig heavy chain V region pe20 - mouse (fragment)  
C:Species: Mus musculus (house mouse)

C;Accession: S55536  
R;Boettger, V.; Boettger, A.; Lane, E.B.; Spruce, B.A.  
T;Mammals; 1947; 1948; 1949; 1950; 1951; 1952; 1953; 1954; 1955; 1956; 1957; 1958; 1959; 1960; 1961; 1962; 1963; 1964; 1965; 1966; 1967; 1968; 1969; 1970; 1971; 1972; 1973; 1974; 1975; 1976; 1977; 1978; 1979; 1980; 1981; 1982; 1983; 1984; 1985; 1986; 1987; 1988; 1989; 1990; 1991; 1992; 1993; 1994; 1995; 1996; 1997; 1998; 1999; 2000; 2001; 2002; 2003; 2004; 2005; 2006; 2007; 2008; 2009; 2010; 2011; 2012; 2013; 2014; 2015; 2016; 2017; 2018; 2019; 2020; 2021; 2022; 2023; 2024; 2025; 2026; 2027; 2028; 2029; 2030; 2031; 2032; 2033; 2034; 2035; 2036; 2037; 2038; 2039; 2040; 2041; 2042; 2043; 2044; 2045; 2046; 2047; 2048; 2049; 2050; 2051; 2052; 2053; 2054; 2055; 2056; 2057; 2058; 2059; 2060; 2061; 2062; 2063; 2064; 2065; 2066; 2067; 2068; 2069; 2070; 2071; 2072; 2073; 2074; 2075; 2076; 2077; 2078; 2079; 2080; 2081; 2082; 2083; 2084; 2085; 2086; 2087; 2088; 2089; 2090; 2091; 2092; 2093; 2094; 2095; 2096; 2097; 2098; 2099; 2100; 2101; 2102; 2103; 2104; 2105; 2106; 2107; 2108; 2109; 2110; 2111; 2112; 2113; 2114; 2115; 2116; 2117; 2118; 2119; 2120; 2121; 2122; 2123; 2124; 2125; 2126; 2127; 2128; 2129; 2130; 2131; 2132; 2133; 2134; 2135; 2136; 2137; 2138; 2139; 2140; 2141; 2142; 2143; 2144; 2145; 2146; 2147; 2148; 2149; 2150; 2151; 2152; 2153; 2154; 2155; 2156; 2157; 2158; 2159; 2160; 2161; 2162; 2163; 2164; 2165; 2166; 2167; 2168; 2169; 2170; 2171; 2172; 2173; 2174; 2175; 2176; 2177; 2178; 2179; 2180; 2181; 2182; 2183; 2184; 2185; 2186; 2187; 2188; 2189; 2190; 2191; 2192; 2193; 2194; 2195; 2196; 2197; 2198; 2199; 2200; 2201; 2202; 2203; 2204; 2205; 2206; 2207; 2208; 2209; 2210; 2211; 2212; 2213; 2214; 2215; 2216; 2217; 2218; 2219; 2220; 2221; 2222; 2223; 2224; 2225; 2226; 2227; 2228; 2229; 2230; 2231; 2232; 2233; 2234; 2235; 2236; 2237; 2238; 2239; 2240; 2241; 2242; 2243; 2244; 2245; 2246; 2247; 2248; 2249; 2250; 2251; 2252; 2253; 2254; 2255; 2256; 2257; 2258; 2259; 2260; 2261; 2262; 2263; 2264; 2265; 2266; 2267; 2268; 2269; 2270; 2271; 2272; 2273; 2274; 2275; 2276; 2277; 2278; 2279; 2280; 2281; 2282; 2283; 2284; 2285; 2286; 2287; 2288; 2289; 2290; 2291; 2292; 2293; 2294; 2295; 2296; 2297; 2298; 2299; 2300; 2301; 2302; 2303; 2304; 2305; 2306; 2307; 2308; 2309; 2310; 2311; 2312; 2313; 2314; 2315; 2316; 2317; 2318; 2319; 2320; 2321; 2322; 2323; 2324; 2325; 2326; 2327; 2328; 2329; 2330; 2331; 2332; 2333; 2334; 2335; 2336; 2337; 2338; 2339; 2340; 2341; 2342; 2343; 2344; 2345; 2346; 2347; 2348; 2349; 2350; 2351; 2352; 2353; 2354; 2355; 2356; 2357; 2358; 2359; 2360; 2361; 2362; 2363; 2364; 2365; 2366; 2367; 2368; 2369; 2370; 2371; 2372; 2373; 2374; 2375; 2376; 2377; 2378; 2379; 2380; 2381; 2382; 2383; 2384; 2385; 2386; 2387; 2388; 2389; 2390; 2391; 2392; 2393; 2394; 2395; 2396; 2397; 2398; 2399; 2400; 2401; 2402; 2403; 2404; 2405; 2406; 2407; 2408; 2409; 2410; 2411; 2412; 2413; 2414; 2415; 2416; 2417; 2418; 2419; 2420; 2421; 2422; 2423; 2424; 2425; 2426; 2427; 2428; 2429; 2430; 2431; 2432; 2433; 2434; 2435; 2436; 2437; 2438; 2439; 2440; 2441; 2442; 2443; 2444; 2445; 2446; 2447; 2448; 2449; 2450; 2451; 2452; 2453; 2454; 2455; 2456; 2457; 2458; 2459; 2460; 2461; 2462; 2463; 2464; 2465; 2466; 2467; 2468; 2469; 2470; 2471; 2472; 2473; 2474; 2475; 2476; 2477; 2478; 2479; 2480; 2481; 2482; 2483; 2484; 2485; 2486; 2487; 2488; 2489; 2490; 2491; 2492; 2493; 2494; 2495; 2496; 2497; 2498; 2499; 2500; 2501; 2502; 2503; 2504; 2505; 2506; 2507; 2508; 2509; 2510; 2511; 2512; 2513; 2514; 2515; 2516; 2517; 2518; 2519; 2520; 2521; 2522; 2523; 2524; 2525; 2526; 2527; 2528; 2529; 2530; 2531; 2532; 2533; 2534; 2535; 2536; 2537; 2538; 2539; 2540; 2541; 2542; 2543; 2544; 2545; 2546; 2547; 2548; 2549; 2550; 2551; 2552; 2553; 2554; 2555; 2556; 2557; 2558; 2559; 2560; 2561; 2562; 2563; 2564; 2565; 2566; 2567; 2568; 2569; 2570; 2571; 2572; 2573; 2574; 2575; 2576; 2577; 2578; 2579; 2580; 2581; 2582; 2583; 2584; 2585; 2586; 2587; 2588; 2589; 2590; 2591; 2592; 2593; 2594; 2595; 2596; 2597; 2598; 2599; 2600; 2601; 2602; 2603; 2604; 2605; 2606; 2607; 2608; 2609; 2610; 2611; 2612; 2613; 2614; 2615; 2616; 2617; 2618; 2619; 2620; 2621

A;Title: Comprehensive epitope analysis of monoclonal anti-proenkephalin antibodies using mutations in the variable region gene.

A;Accession: S55536  
A;Status: preliminary  
A;Molecular type: RNA

A:Residues: 1-120 <BOE>  
A:Cross-references: EMBL:X82589; NID:g854304; PIDN:CAAS7925.1; PID:g854305  
A:Suppl:1: Immunoglobulin Y reagent immunoglobulin Y reagent

**Keywords:** heterotetramer; immunoglobulin F<sub>14-97</sub>/Domain: immunoglobulin homology <IMM>

Query Match	58.4%	Score 396.5;	DB 2,	Length 120;
Best Local Similarity	67.5%;	Pred. No. 7.4e-28;		
Matches	81	Conserved	11	Mismatches 37
				Indels 1
				Gaps 1

2 VOIOBGGGLVAGGSLRLSCAAGTIIITIMDMYRQTPGKORELVGRTAGGS-TNYA 60

Db 1V0L0ESGGGLVWPGSGSLKLSCAAGGFPTSTVYAMSVMWQREKLEWATISSGGSXYTF 60

Qy 61 DAKKERFTISDNNMNTYTLQMSLKEPEDAVYVYCNALLIRRMKSKSNDVWGGGTQVYSS 120

61 DSVKGRFTISRQNAKNTLYLQMSLKEPEDAVYVYCARLYYDVPYMDWVGQITVYSS 120

## RESULT 2

## Ig heavy

CiDate: 21-Dec-1990 #sequence\_revision 21-Dec-1990 #text\_change 16-Dec-1998  
CiAccession: H36005

Proc. Natl. Acad. Sci. U.S.A. 87, 6146-6150, 1990

A;Accession: H36005  
A;Status: preliminary

A;Residues: 1-121 <SCH>  
A;Cross-references: GB:M34032

A;Gene: GDB:IGH@;IGHDY1  
A;Cross-references: GDB:118731; OMIM:146910

C:Superfamily: immunoglobulin V region; immunoglobulin homology  
 C:Keywords: heterotetramer; immunoglobulin  
 F:15-100/Domain: immunoglobulin homology <IMM>

Query Match 57.8%; Score 392.5; DB 2; Length 121;  
 Best Local Similarity 66.7%; Pred. No. 1.7e-27;  
 Matches 82; Conservative 10; Mismatches 26; Indels 5; Gaps 2;

QY 1 VOLOESGGGLVQAGSLRLSCAAGTILSIITMDYRQTPGKRELVGRI---TAGSGT 57  
 DB 1 EVQLVSGGGLVQPGGSLRLSCAAGTFTPSYAMMWVRQAPGKGLFWGRIKSKTDGQT 60

QY 58 NVADAKGRFTISKDNKNTVYLQNMSLKPEDTAVYYCNALITRMDKSVNDYWGQGTQVT 117  
 DB 61 DVAAPKGRFTISRDSKNTLYLQNMSLKTEDTAVYYCTT--DRGSSGGDWGGTILVT 118

QY 118 VSS 120  
 DB 119 VSS 121

## RESULT 3

S3114  
 Ig heavy chain - human

C:Species: Homo sapiens (man)  
 C/Date: 02-Dec-1993 #sequence\_revision 26-May-1995 #text\_change 17-Mar-1999

C/Accession: S31114  
 R:Raphoport, F.M.; Timmers, E.; Kenter, M.J.H.; van Tol, M.J.D.; Vossen, J.M.; Schuurman  
 Eur. J. Immunol. 22, 247-251, 1992

A/Title: Restricted utilization of germ-line V(H)3 genes and short diverse third complement  
 A/Reference number: S31104; PMID:92111633; PMID:1130252

A/Accession: S31114  
 A/Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A/Molecule type: mRNA

A/Residues: 1-123 <RAA>

A/Cross-references: EMBL:X62963  
 A/Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1991

C:Superfamily: immunoglobulin V region; immunoglobulin homology  
 C:Keywords: heterotetramer; immunoglobulin  
 F:15-98/Domain: immunoglobulin homology <IMM>

Query Match 57.7%; Score 391.5; DB 2; Length 123;  
 Best Local Similarity 67.5%; Pred. No. 2.1e-27;  
 Matches 83; Conservative 11; Mismatches 26; Indels 3; Gaps 2;

QY 1 QVQLQESGGGLVQAGSLRLSCAAGTILSIITMDYRQTPGKRELVGRI--AGGSTY 59  
 DB 1 EVQLVSGGGLVQPGGSLRLSCAAGTFTPSYAMMWVRQAPGKGLFWSAISGSGSTY 60

QY 60 ADSAKGRFTISKDNKNTVYLQNMSLKPEDTAVYYC--NALITRMDKSVNDYWGQGTQVT 117  
 DB 61 ADSVKGRTISRDSKNTLYLQNMSLRAEDTAVYYCAKASLVLRFLWLPDYWGQGTILVT 120

QY 118 VSS 120  
 DB 121 VSS 123

## RESULT 4

S55537  
 Ig heavy chain V region pe21 - mouse (fragment)

C:Species: Mus musculus (house mouse)  
 C/Date: 27-Oct-1995 #sequence\_revision 03-Nov-1995 #text\_change 23-Jul-1999

C/Accession: S55537  
 R:Boettger, V.; Boettger, A.; Lane, E.B.; Spruce, B.A.  
 J. Mol. Biol. 247, 932-946, 1995

A/Title: Comprehensive epitope analysis of monoclonal anti-proenkephalin antibodies using  
 mutations in the variable region genes.  
 A/Reference number: S55528; PMID:95239763; PMID:7536850

A/Accession: S55537  
 A/Status: preliminary

A/Molecule type: mRNA  
 A/Residues: 1-120 <BOE>

A/Cross-references: EMBL:X62590; NID:g854306; PIDD:CA457926.1; PID:g854307  
 C:Superfamily: immunoglobulin V region; immunoglobulin homology  
 C:Keywords: heterotetramer; immunoglobulin  
 F:14-97/Domain: immunoglobulin homology <IMM>

Query Match 57.4%; Score 389.5; DB 2; Length 120;  
 Best Local Similarity 66.7%; Pred. No. 3e-27;  
 Matches 80; Conservative 11; Mismatches 28; Indels 1; Gaps 1;

QY 2 VOLOESGGGLVQAGSLRLSCAAGTILSIITMDYRQTPGKRELVGRI--TAGSGT 60  
 DB 1 VOQLVSGGGLVQPGGSLRLSCAAGTFTPSYAMMWVRQAPGKGLFWGRIKSKTDGQT 60

QY 61 DSAKGRFTISKDNKNTVYLQNMSLKPEDTAVYYCNALITRMDKSVNDYWGQGTQVT 120  
 DB 61 DSVKGRFTISRDSKNTLYLQNMSLKPEDTAVYYCTRLVYDDPYMDYWGQGTILVT 120

## RESULT 5

S26798  
 Ig heavy chain V region - human

C:Species: Homo sapiens (man)  
 C/Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 20-Jun-2000

C/Accession: S26798  
 R:Mortari, F.; Newton, J.A.; Wang, J.Y.; Schroeder Jr., H.W.  
 Eur. J. Immunol. 22, 241-245, 1992

A/Title: The human cord blood antibody repertoire. Frequent usage of the V(H)7 gene family  
 A/Reference number: S26786; PMID:92111632; PMID:1130251

A/Accession: S26798  
 A/Status: preliminary  
 A/Molecule type: mRNA

A/Residues: 1-121 <MOR>  
 A/Cross-references: EMBL:X61015; NID:932795; PIDD:CAA43349.1; PID:91335126

C:Superfamily: immunoglobulin V region; immunoglobulin homology  
 C:Keywords: heterotetramer; immunoglobulin  
 F:15-98/Domain: immunoglobulin homology <IMM>

Query Match 57.4%; Score 389.5; DB 2; Length 121;  
 Best Local Similarity 66.9%; Pred. No. 3.1e-27;  
 Matches 81; Conservative 8; Mismatches 31; Indels 1; Gaps 1;

QY 1 QVQLQESGGGLVQAGSLRLSCAAGTILSIITMDYRQTPGKRELVGRI--TAGSGT 59  
 DB 1 EVQLVSGGGLVQPGGSLRLSCAAGTFTPSYAMMWVRQAPGKGLFWSHIKKGGITTY 60

QY 60 ADSAKGRFTISKDNKNTVYLQNMSLKPEDTAVYYCNALITRMDKSVNDYWGQGTQVT 119  
 DB 61 ADSVKGRTISRDSKNTLYLQNMSLRGEDTAVYYCVRAIGRYSYGLDYWGQGTILVT 120

QY 120 S 120  
 DB 121 S 121

## RESULT 6

A36006  
 Ig heavy chain V region (M26) - human

C:Species: Homo sapiens (man)  
 C/Date: 21-Dec-1990 #sequence\_revision 21-Dec-1990 #text\_change 16-Dec-1998

C/Accession: A36006  
 R:Schroeder Jr., H.W.; Wang, J.Y.  
 Proc. Natl. Acad. Sci. U.S.A. 87, 6146-6150, 1990

A/Title: Preferential utilization of conserved immunoglobulin heavy chain variable gene  
 A/Reference number: A36005; PMID:90349571; PMID:2117273

A/Accession: A36006  
 A/Status: preliminary

A/Molecule type: mRNA  
 A/Residues: 1-123 <SCH>

A/Cross-references: GB:M34023  
 C:Genetics:  
 A/Gene: GDB:IGH@; IGHDI1  
 A/Cross-references: GDB:118731; OMIM:146910

A/Map position: 14q32.33-14q32.33

C:Superfamily: immunoglobulin V region; immunoglobulin homology  
C:Keywords: heterotetramer; immunoglobulin  
F.15-100/Domain: immunoglobulin homology <IMM>

Query Match 57.4%; Score 389.5; DB 2; Length 123;  
Best Local Similarity 65.9%; Pred. No. 3.1e-27;  
Matches 81; Conservative 10; Mismatches 25; Indels 3; Gaps 1;

QY 1 QVQLQSGGGLVQAGSRLSCAASGTLISITIMDWRQTPGKORLVRIT---TAGGST 57  
DB 1 EVQLVESGGGLVQPGGSLRLSCAASGTFEFNAMWVRQAPGKLEWVGRIKSKTDGTT 60  
QY 58 NYADAKRFTISKDNKNTVYLQNNSLKPEDTAVYCNALITRDKSVNDYMGQGTQVT 117  
DB 61 DYAPVKRFTISRDNKNTLYLQNNSLKTEDTAVYCTTSIPGIYAVAGTVMGQGLT 120  
QY 118 VSS 120  
DB 121 VSS 123

## RESULT 7

S31110  
Ig heavy chain - human  
C:Species: Homo sapiens (man)  
C>Date: 02-Dec-1993 #sequence\_revision 26-May-1995 #text\_change 17-Mar-1999  
C/Accession: S31110  
R:Raaphorst, F.M.; Timmers, E.; Kenter, M.J.H.; van Tol, M.J.D.; Vossen, J.M.; Schuurman  
Bur, J. Immunol. 22, 247-251, 1992

A:Title: Restricted utilization of germ-line V(H)3 genes and short diverse third complement  
A:Reference number: S31104; MUID:92111633; PMID:1730252  
A:Accession: S31110  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: mRNA  
A:Residues: 1-116 <RAA>  
A:Cross-references: EMBL:X62958  
A>Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1991  
C:Superfamily: immunoglobulin V region; immunoglobulin homology  
C:Keywords: heterotetramer; immunoglobulin  
F.15-97/Domain: immunoglobulin homology <IMM>

Query Match 57.3%; Score 389; DB 2; Length 116;  
Best Local Similarity 66.7%; Pred. No. 3.2e-27;  
Matches 80; Conservative 10; Mismatches 26; Indels 4; Gaps 1;

QY 1 QVQLQSGGGLVQAGSRLSCAASGTLISITIMDWRQTPGKORLVRITAGGSTNYA 60  
DB 1 EVQLVESGGGLVQPGGSLRLSCAASGTFVSSVMWVRQAPGKLEWVSIVYSGSTIYA 60  
QY 61 DSAKGRFTISKDNKNTVYLQNNSLKPEDTAVYCNALITRDKSVNDYMGQGTQVT 120  
DB 61 DSVKGRFTISRDNKNTLYLRNLSLRADTAVYVC---ARDLGGGGDYMGQGLTVTSS 116

## RESULT 8

C36005  
Ig heavy chain V region (30p1) - human  
C:Species: Homo sapiens (man)  
C>Date: 21-Dec-1990 #sequence\_revision 21-Dec-1990 #text\_change 16-Aug-1996  
C/Accession: C36005  
R:Schroeder Jr., H.W.; Mang, J.Y.  
Proc. Natl. Acad. Sci. U.S.A. 87, 6146-6150, 1990  
A:Title: Preferential utilization of conserved immunoglobulin heavy chain variable gene  
A:Reference number: A6005; MUID:90349571; PMID:2117273  
A:Accession: C36005  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-119 <SCH>

A:Cross-references: GB:M18513  
C:Superfamily: immunoglobulin V region; immunoglobulin homology  
C:Keywords: heterotetramer; immunoglobulin  
F.15-98/Domain: immunoglobulin homology <IMM>

Query Match 57.2%; Score 388.5; DB 2; Length 119;  
Best Local Similarity 68.6%; Pred. No. 3.7e-27;  
Matches 83; Conservative 7; Mismatches 28; Indels 3; Gaps 3;

QY 1 QVQLQSGGGLVQAGSRLSCAASGTLISITIMDWRQTPGKORLVRITAGGSTNY 59  
DB 1 EVQLVESGGGLVQPGGSLRLSCAASGTFSSVMWVRQAPGKLEWVAISGSGSTIY 60  
QY 60 ADSAKRFTISKDNKNTVYLQNNSLKPEDTAVYCNALITRDKSVNDYMGQGTQVT 119  
DB 61 ADSVGRFTISRDNKNTLYLQNNSLRADTAVYVC-ADAGGSGSF-DYMGQGLTVTS 118  
QY 120 S 120  
DB 119 S 119

## RESULT 9

S48798  
Ig heavy chain V region (anti-Sm, VH3/Dxp4/JH4b) - human  
C:Species: Homo sapiens (man)  
C>Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 23-Jul-1999  
C/Accession: S48798  
R:Mahmoudi, M.; Edwards, J.; Cairns, E.; Bell, D.  
Submitted to the EMBL Data Library, October 1994  
A:Description: Molecular characterization of natural human anti-Sm autoantibodies.  
A:Reference number: S48797  
A:Accession: S48798  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-120 <MAH>

A:Cross-references: EMBL:Z46382; NID:9562324; PIDN:CA86621.1; PID:g1340167  
C:Superfamily: immunoglobulin V region; immunoglobulin homology  
C:Keywords: heterotetramer; immunoglobulin  
F.15-98/Domain: immunoglobulin homology <IMM>

Query Match 57.1%; Score 388; DB 2; Length 120;  
Best Local Similarity 67.8%; Pred. No. 4.1e-27;  
Matches 82; Conservative 7; Mismatches 30; Indels 2; Gaps 2;

QY 1 QVQLQSGGGLVQAGSRLSCAASGTLISITIMDWRQTPGKORLVRITAGGSTNY 59  
DB 1 EVQLVESGGGLVQPGGSLRLSCAASGTFSSVMWVRQAPGKLEWVAISGSGSTIY 60  
QY 60 ADSAKRFTISKDNKNTVYLQNNSLKPEDTAVYCNALITRDKSVNDYMGQGTQVT 119  
DB 61 ADSVGRFTISRDNKNTLYLQNNSLRADTAVYCAKGRGFW-SGYKDYMGQGLTVTS 119  
QY 120 S 120  
DB 120 S 120

## RESULT 10

I55673  
Ig heavy chain - human (fragment)  
C:Species: Homo sapiens (man)  
C>Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 21-Jan-2000  
C/Accession: I55673  
R:Knight, G.B.; Agnello, V.; Bongura, V.; Barnes, J.L.; Parke, D.J.; Zhang, Q.X.  
U. Exp. Med. 178, 1503-1511, 1993  
A:Title: Human rheumatoid factor cross-idiotypes. IV. Studies on WA kid-positive IgM with  
clint from the 17.109 and G6 Xids.  
A:Reference number: I55673; MUID:94055558; PMID:8245772  
A:Accession: I55673  
A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA  
A:Residues: 1-121 <RES>  
A:Cross-references: GB:M87268; NID:g186197; PIDN:AAC37536.1; PID:g186198  
A:Gene: GDB:IGHM  
A:Cross-references: GDB:120086; OMIM:147020  
A:Map position: 14q32.33-14q32.33

C:Superfamily: immunoglobulin V region; immunoglobulin homology  
F:15-98/Domain: immunoglobulin homology <IMM>

Query Match 57.1%; Score 387.5; DB 2; Length 121;  
Best Local Similarity 66.9%; Pred. No. 4.6e-27;  
Matches 81; Conservative 7; Mismatches 32; Indels 1; Gaps 1;

QY 1 QVQLQESGGGLVQAGSLRLSCAASGTTISITIMMYRQTPGKRELVGRIT-AGGSITNY 59  
1 EVQLVESGGGLVQPGGSLRLSCAASGTFSTYGVGMWRAPGKGLEWVSALISGSGSTYY 60  
DB 60 ADSAKGRFTISKDNAAKNTVYLQNSLKPEPTAVYYCNALITRMDKSVNDYWGQGTQVTVS 119  
61 ADSVKGKFTISRDNKNTLYLQNSLRADTAIVYCAAPRAAGSPYPDYWGQGTIVTVSS 120  
QY 120 S 120  
DB 121 S 121

## RESULT 11

S55539  
Ig heavy chain V region pe24 - mouse (fragment)  
C:Species: Mus musculus (house mouse)  
C:Date: 27-Oct-1995 #sequence\_revision 03-Nov-1995 #text\_change 23-Jul-1999  
C:Accession: S55539  
R:Boettger, V.; Boettger, A.; Lane, E.B.; Spruce, B.A.  
J. Mol. Biol. 247, 932-946, 1995  
A:Title: Comprehensive epitope analysis of monoclonal anti-prenkephalin antibodies using  
utations in the variable region genes.  
A:Reference number: S55528; MUID:95239763; PMID:7536650  
A:Accession: S55539  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-120 <BOE>  
A:Cross-references: EMBL:X82593; NID:9854312; PIDN:CA57929.1; PID:9854313  
C:Superfamily: immunoglobulin V region; immunoglobulin homology  
C:Keywords: heterotrimer; immunoglobulin  
F:14-97/Domain: immunoglobulin homology <IMM>

Query Match 56.8%; Score 385.5; DB 2; Length 120;  
Best Local Similarity 65.8%; Pred. No. 6.8e-27;  
Matches 79; Conservative 12; Mismatches 28; Indels 1; Gaps 1;  
QY 2 VQLQESGGGLVQAGSLRLSCAASGTTISITIMMYRQTPGKRELVGRIT-AGGS-ITNYA 60  
1 VQLQESGGGLVQPGGSLRLSCAASGTFSTYGVGMWRAPGKLEWVSALISGSGSTYY 60  
DB 1 VQLQESGGGLVQPGGSLRLSCAASGTFSTYGVGMWRAPGKLEWVSALISGSGSTYY 60  
QY 61 ADSAKGRFTISKDNAAKNTVYLQNSLKPEPTAVYYCNALITRMDKSVNDYWGQGTQVTVSS 120  
61 ADSVKGKFTISRDNKNTLYLQNSLRADTAIVYCAAPRAAGSPYPDYWGQGTIVTVSS 120  
DB 61 ADSVKGKFTISRDNKNTLYLQNSLRADTAIVYCAAPRAAGSPYPDYWGQGTIVTVSS 120

## RESULT 12

S31109  
Ig heavy chain - human  
C:Species: Homo sapiens (man)  
C:Date: 02-Dec-1993 #sequence\_revision 26-May-1995 #text\_change 17-Mar-1999  
C:Accession: S31109  
R:Raaphorst, F.M.; Timmers, E.; Kenter, M.J.H.; van Tol, M.J.D.; Vossen, J.M.; Schuurman,  
Eur. J. Immunol. 22, 247-251, 1992  
A:Title: Restricted utilization of germ-line V(H)3 genes and short diverse third comple  
A:Reference number: S31104; MUID:92111633; PMID:1730252  
A:Accession: S31109  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: mRNA  
A:Residues: 1-117 <RAA>  
A:Cross-references: EMBL:X62960  
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1991  
C:Superfamily: immunoglobulin V region; immunoglobulin homology  
C:Keywords: heterotrimer; immunoglobulin  
F:15-100/Domain: immunoglobulin homology <IMM>

Query Match 56.6%; Score 384.5; DB 2; Length 117;  
Best Local Similarity 65.9%; Pred. No. 8.1e-27;  
Matches 81; Conservative 9; Mismatches 24; Indels 9; Gaps 2;

QY 1 QVQLQESGGGLVQAGSLRLSCAASGTTISITIMMYRQTPGKRELVGRIT-AGGSIT 57  
1 EVQLVESGGGLVQPGGSLRLSCAASGTFSTYGVGMWRAPGKGLEWVSALISGSGSTYY 60  
DB 1 EVQLVESGGGLVQPGGSLRLSCAASGTFSTYGVGMWRAPGKGLEWVSALISGSGSTYY 60  
QY 58 NYADSAKGRFTISKDNAAKNTVYLQNSLKPEPTAVYYCNALITRMDKSVNDYWGQGTQV 117  
61 NYADPAKGRFTISRDNKNTLYLQNSLKPEPTAVYYCAAT- - - - -YFDYWGQGTIVTV 114  
DB 118 VSS 120  
115 VSS 117

## RESULT 13

S31108  
Ig heavy chain - human  
C:Species: Homo sapiens (man)  
C:Date: 02-Dec-1993 #sequence\_revision 26-May-1995 #text\_change 17-Mar-1999  
C:Accession: S31108  
R:Raaphorst, F.M.; Timmers, E.; Kenter, M.J.H.; van Tol, M.J.D.; Vossen, J.M.; Schuurman,  
Eur. J. Immunol. 22, 247-251, 1992  
A:Title: Restricted utilization of germ-line V(H)3 genes and short diverse third comple  
A:Reference number: S31104; MUID:92111633; PMID:1730252  
A:Accession: S31108  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: mRNA  
A:Residues: 1-119 <RAA>  
A:Cross-references: EMBL:X62956  
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1991  
C:Superfamily: immunoglobulin V region; immunoglobulin homology  
C:Keywords: heterotrimer; immunoglobulin  
F:15-98/Domain: immunoglobulin homology <IMM>

Query Match 56.6%; Score 384.5; DB 2; Length 119;  
Best Local Similarity 66.1%; Pred. No. 8.2e-27;  
Matches 82; Conservative 9; Mismatches 24; Indels 9; Gaps 3;  
QY 1 QVQLQESGGGLVQAGSLRLSCAASGTTISITIMMYRQTPGKRELVGRIT-AGGSITNY 59  
1 EVQLVESGGGLVQPGGSLRLSCAASGTFSTYGVGMWRAPGKLEWVSALISGSGSTYY 60  
DB 1 EVQLVESGGGLVQPGGSLRLSCAASGTFSTYGVGMWRAPGKLEWVSALISGSGSTYY 60  
QY 60 ADSAKGRFTISKDNAAKNTVYLQNSLKPEPTAVYYCNALITRMDKSVN- - -DYWGQGTQV 116  
61 ADSVKGKFTISRDNKNTLYLQNSLRADTAIVYCAK- - - - -DRRLTGTFDIWGQGTIV 115  
DB 61 ADSVKGKFTISRDNKNTLYLQNSLRADTAIVYCAK- - - - -DRRLTGTFDIWGQGTIV 115  
QY 117 TVSS 120  
116 TVSS 119

## RESULT 14

S31120  
Ig heavy chain - human  
C:Species: Homo sapiens (man)  
C:Date: 02-Dec-1993 #sequence\_revision 26-May-1995 #text\_change 17-Mar-1999  
C:Accession: S31120  
R:Raaphorst, F.M.; Timmers, E.; Kenter, M.J.H.; van Tol, M.J.D.; Vossen, J.M.; Schuurman,  
Eur. J. Immunol. 22, 247-251, 1992  
A:Title: Restricted utilization of germ-line V(H)3 genes and short diverse third comple  
A:Reference number: S31104; MUID:92111633; PMID:1730252  
A:Accession: S31120  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: mRNA  
A:Residues: 1-114 <RAA>  
A:Cross-references: EMBL:X62972  
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1991  
C:Superfamily: immunoglobulin V region; immunoglobulin homology  
C:Keywords: heterotrimer; immunoglobulin  
F:15-98/Domain: immunoglobulin homology <IMM>



This Page Blank (uspto)



Tue Oct 5 09:51:13 2004

us-09-805-290a-20.rsp

Page 1

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2004, 07:29:35 / Search time 11.5463 Seconds  
(without alignments)  
581.749 Million cell updates/sec

Title: US-09-805-290A-20

Perfect score: 679  
Sequence: 1 QVQLQESGGGLVQAGSLRLT.....WGQGTQVTVSSPEKTPKPP 129

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 141681 segs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_42:\*

Pred. NO. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	351	51.7	136 1 HV16_MOUSE	P01783 mus musculu
2	350	51.5	116 1 HV05_CARAU	P19181 carassius a
3	349	51.4	122 1 HV3A_HUMAN	P01762 homo sapien
4	347.5	51.2	115 1 HV3F_HUMAN	P01767 homo sapien
5	341	50.2	116 1 HV3T_HUMAN	P01781 homo sapien
6	340	50.1	126 1 HV3K_HUMAN	P01772 homo sapien
7	337.5	49.7	119 1 HV3L_HUMAN	P01773 homo sapien
8	334.5	49.3	117 1 HV3C_HUMAN	P01764 homo sapien
9	334	49.2	111 1 HV02_MOUSE	P01804 mus musculu
10	333.5	49.1	117 1 HV02_CARAU	P01785 canis fami
11	333.5	49.1	118 1 HV39_MOUSE	P01809 mus musculu
12	333	49.0	120 1 HV3U_HUMAN	P01782 homo sapien
13	330	48.6	114 1 HV01_CARAU	P01784 canis fami
14	330	48.6	114 1 HV3B_HUMAN	P01787 homo sapien
15	330	48.6	119 1 HV37_MOUSE	P01807 mus musculu
16	330	48.6	122 1 HV3G_HUMAN	P01768 homo sapien
17	329.5	48.5	117 1 HV53_MOUSE	P18524 mus musculu
18	329	48.5	119 1 HV40_MOUSE	P01810 mus musculu
19	327.5	48.2	115 1 HV33_MOUSE	P01802 mus musculu
20	326.5	48.1	115 1 HV3D_HUMAN	P01765 homo sapien
21	326	48.0	119 1 HV38_MOUSE	P01808 mus musculu
22	324.5	47.8	121 1 HV3J_HUMAN	P01771 homo sapien
23	323.5	47.6	113 1 HV32_MOUSE	P01799 mus musculu
24	323.5	47.6	115 1 HV32_MOUSE	P01801 mus musculu
25	323.5	47.6	120 1 HV57_MOUSE	P01766 homo sapien
26	321.5	47.3	98 1 HV57_MOUSE	P18528 mus musculu
27	321.5	47.3	119 1 HV3I_HUMAN	P01770 homo sapien
28	319.5	47.1	116 1 HV3R_HUMAN	P01779 homo sapien
29	319.5	47.1	117 1 HV41_MOUSE	P01811 mus musculu
30	319.5	47.1	117 1 HV55_MOUSE	P18526 mus musculu
31	319	47.0	97 1 HV56_MOUSE	P18527 mus musculu
32	319	47.0	122 1 HV3H_HUMAN	P01769 homo sapien
33	318.5	46.9	113 1 HV27_MOUSE	P01796 mus musculu

34	317.5	46.8	117 1 HV42_MOUSE	P01812 mus musculu
35	316	46.5	113 1 HV3P_HUMAN	P01777 homo sapien
36	314.5	46.3	119 1 HV31_MOUSE	P01800 mus musculu
37	314.5	46.3	117 1 HV54_MOUSE	P18525 mus musculu
38	314	46.2	122 1 HV20_MOUSE	P01789 mus musculu
39	313.5	46.2	117 1 HV17_MOUSE	P01786 mus musculu
40	312.5	46.0	113 1 HV28_MOUSE	P01797 mus musculu
41	310.5	45.7	123 1 HV24_MOUSE	P01793 mus musculu
42	309	45.5	115 1 HV38_HUMAN	P01780 homo sapien
43	308.5	45.4	113 1 HV29_MOUSE	P01798 mus musculu
44	307.5	45.3	123 1 HV22_MOUSE	P01791 mus musculu
45	307.5	45.3	144 1 HV26_MOUSE	P01795 mus musculu

#### ALIGNMENTS

```
RESULT 1
HV16_MOUSE          STANDARD;          PRT;          136 AA.
ID HV16_MOUSE
AC P01783;
DT 21-JUL-1986 (Rel. 01, Created)
DI 21-JUL-1986 (Rel. 01, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE 1g heavy chain V region MOPC 21 precursor (Fragment).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OX Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN NCBI_Taxid=10090;
RP [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=61234548; PubMed=6788376;
RA Botheil A.L.M., Paskind M., Reih M., Imanishi-Kari T., Rajewsky K.,
RA Baltimore D.;
RT "Heavy chain variable region contribution to the NPb family of
RT antibodies: somatic mutation evident in a gamma 2a variable region.";
RL Cell 24:625-637(1981).
RN [2]
RP SEQUENCE OF 17-136.
RX MEDLINE=77100368; PubMed=401950;
RA Aetugbo K., Milstein C., Secher D.S.;
RT "Molecular analysis of spontaneous somatic mutants.";
RL Nature 285:299-304(1977).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-stb.ch).
CC -----
DR EMBL; J00522; AAD15290.1; -.
DR PIR; E90809; GIMS21.
DR PDB; 1IGC; 03-JUN-95.
DR InterPro; IPR007110; 1g-like.
DR InterPro; IPR003596; 1g_V.
DR Pfam; PF00047; 1g_V.
DR SMART; SM00406; 1g_V.
DR PROSITE; PS50835; 1g_LIKE; 1.
KW Immunoglobulin V region; Signal; 3D-structure.
FT NON TER 1
FT SIGNAL 1
FT CHAIN 17 136 IG HEAVY CHAIN V REGION MOPC 21.
FT DOMAIN 115 119 D SEGMENT.
FT DOMAIN 120 136 JH4 SEGMENT.
FT DISULFID 38 112
FT CONFLICT 75 78 HVAD -> DYAH (IN REF. 2).
FT CONFLICT 89 90 DN -> ND (IN REF. 2).
FT CONFLICT 115 115 W -> H (IN REF. 2).
FT CONFLICT 120 120 Y -> W (IN REF. 2).
FT NON TER 136
SQ SEQUENCE 136 AA; 15071 MW; 2276A98DBDBF7016 CRC64;
```



CC REGION OF THIS MYELOMA PROTEIN IS ALSO GIVEN.  
 CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.  
 DR PIR: A02050; A2HUBU.  
 DR HSSP; P01789; 1MCP.  
 DR GO: GO:0005576; C:extracellular; NAS.  
 DR GO: GO:0003823; F:antigen binding; NAS.  
 DR GO: GO:0006955; P:immune response; NAS.  
 DR InterPro; IPR007110; IG-like.  
 DR InterPro; IPR003596; IG\_v.  
 DR Pfam; PF00047; Ig; 1.  
 DR SMART; SM00406; IGV; 1.  
 DR PROSITE; PS50835; IG LIKE; 1.  
 DR Immunoglobulin V region.  
 FT DOMAIN 1 111 IG-LIKE.  
 FT NON\_TER 115 115  
 SQ SEQUENCE 115 AA; 12379 MW; 208876A7DF52DCFA CRC64;  
 Query Match 51.2%; Score 347.5; DB 1; Length 115;  
 Best Local Similarity 59.2%; Pred. No. 1.4e-26;  
 Matches 71; Conservative 16; Mismatches 28; Indels 5; Gaps 1;  
 QY 1 QVQLQESGGGLVQAGSRLRSCAAGTLLSTIYMDYRQTPKORELVGRITAGSTNYA 60  
 DB 1 EVQLVETGGGLVQPGSRLRSCAAGFTVSHSMWVQAQFGKALZWVSATYRGSTYYA 60  
 QY 61 DSAGKRFITSDNKNNTYILQNSLKPEDTAVYVCNALITWDKSVNDYWGQGTQVTVSS 120  
 DB 61 DSVKGRFTISRDSDRYVLTQMSLRVEDTALVYC---ARGM--GGGDIWGQGTIVTVSS 115  
 RESULT 5  
 HV3K HUMAN STANDARD; PRT; 116 AA.  
 ID HV3K HUMAN  
 AC P01781.  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Ig heavy chain V-II region GAL.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 ON NCBI\_TaxID=9606;  
 RN [1]  
 RP MEDLINE=75059123; PubMed=4803843;  
 RA Watanabe S., Barnikol H.U., Horn J., Berram J., Hilschmann N.;  
 RT "The primary structure of a monoclonal IGM-immunoglobulin  
 (macroglobulin Gal.), II: the amino acid sequence of the H-chain (mu-  
 type), subgroup H II. Architecture of the complete IGM-molecule."  
 RT Hope-Seyler's Z. Physiol. Chem. 354:1505-1509(1973).  
 RL [2]  
 RP REVISION TO 28-33.  
 RA Hilschmann N.;  
 RL Submitted (JUN-1975) to the PIR data bank.  
 CC -1- MISCELLANEOUS: THIS MO CHAIN WAS ISOLATED FROM A WALDENSTROM'S  
 CC MACROGLOBULIN.  
 CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.  
 DR PIR; A02064; M3HUGL.  
 DR HSSP; P01772; 2FB4.  
 DR GO: GO:0005576; C:extracellular; NAS.  
 DR GO: GO:0003823; F:antigen binding; NAS.  
 DR GO: GO:0006955; P:immune response; NAS.  
 DR InterPro; IPR007110; IG-like.  
 DR InterPro; IPR003596; IG\_v.  
 DR Pfam; PF00047; Ig; 1.  
 DR SMART; SM00406; IGV; 1.  
 DR PROSITE; PS50835; IG LIKE; 1.  
 DR Immunoglobulin V region.  
 FT DOMAIN 1 112 IG-LIKE.  
 FT NON\_TER 116 116  
 SQ SEQUENCE 116 AA; 12730 MW; 2C67CA9AAAA1282 CRC64;  
 Query Match 50.2%; Score 341; DB 1; Length 116;

Best Local Similarity 61.2%; Pred. No. 6.1e-26;  
 Matches 74; Conservative 9; Mismatches 32; Indels 6; Gaps 3;  
 QY 1 QVQLQESGGGLVQAGSRLRSCAAGTLLSTIYMDYRQTPKORELVGRITAGGS-TNY 59  
 DB 1 EVQLVSGGDLVQPGSRLRSCAAGFBRBLGKTWRAPKQGLEWVANIKZBSZZBY 60  
 QY 60 ADSAKRFTISKDNANVTYILQNSLKPEDTAVYVCNALITWDKSVNDYWGQGTQVTVS 119  
 DB 61 VDSVKGRTISRDNKNNTYILQNSLRVEDTALVYC---ARGM--GGGDIWGQGTIVTVS 115  
 QY 120 S 120  
 DB 116 T 116  
 RESULT 6  
 HV3K HUMAN STANDARD; PRT; 126 AA.  
 ID HV3K HUMAN  
 AC P01772;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Ig heavy chain V-II region KOL.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 ON NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE, AND DISULFIDE BONDS.  
 RX MEDLINE=83289131; PubMed=6884994;  
 RA Schmidt W.E., Jung H.-D., Palm W., Hilschmann N.;  
 RT "Three-dimensional structure determination of antibodies. Primary  
 structure of crystallized monoclonal immunoglobulin IgG1 KOL, I.";  
 RL Hoppe-Seyler's Z. Physiol. Chem. 364:713-747(1983).  
 RL [2]  
 RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).  
 RX MEDLINE=81072295; PubMed=7441755;  
 RA Marguier M., Delsenhofer J., Huber R., Palm W.;  
 RT "Crystallographic refinement and atomic models of the intact  
 immunoglobulin molecule KOL and its antigen-binding fragment at 3.0 A  
 and 1.0-A resolution."  
 RL J. Mol. Biol. 141:369-391(1980).  
 CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.  
 DR PIR; A02055; G1HUKL.  
 DR PDB; 2FG3; 12-JUL-89.  
 DR GO: GO:0005576; C:extracellular; NAS.  
 DR GO: GO:0003823; F:antigen binding; NAS.  
 DR GO: GO:0006955; P:immune response; NAS.  
 DR InterPro; IPR007110; IG-like.  
 DR InterPro; IPR003596; IG\_v.  
 DR Pfam; PF00047; Ig; 1.  
 DR SMART; SM00406; IGV; 1.  
 DR PROSITE; PS50835; IG LIKE; 1.  
 DR Immunoglobulin V region; 3D-structure; Pyrrolidone carboxylic acid.  
 KW Immunoglobulin V region; 3D-structure; Pyrrolidone carboxylic acid.  
 FT DOMAIN 1 112  
 FT MOD\_RES 1 1  
 FT DISULFID 22 96  
 FT DISULFID 105 110  
 FT STRAND 3 7  
 FT STRAND 11 12  
 FT STRAND 14 15  
 FT TURN 18 25  
 FT STRAND 29 31  
 FT HELIX 34 39  
 FT STRAND 41 42  
 FT TURN 45 51  
 FT TURN 53 54  
 FT STRAND 58 60  
 FT HELIX 62 64  
 FT STRAND 65 65  
 FT TURN 66 67

FT STRAND 68 73  
 FT TURN 74 77  
 FT STRAND 78 83  
 FT HELIX 88 90  
 FT STRAND 92 99  
 FT STRAND 106 106  
 FT TURN 107 108  
 FT STRAND 109 109  
 FT STRAND 113 116  
 FT STRAND 120 124  
 FT NON TER 126 126  
 SQ SEQUENCE 126 AA; 13718 MW; E4D71B52B16F876 CRC64;

Query Match 50.1%; Score 340; DB 1; Length 126;  
 Best Local Similarity 58.7%; Pred. No. 8,4e-26;  
 Matches 74; Conservative 13; Mismatches 33; Indels 6; Gaps 2;

QY 1 QVQLQESGGGLVQAGSLRLSCAASGTLISITIMDMYRQTPGKQRELYGRTITAGGS-TNY 59  
 1 QVQLVSSGGGVQPGSLRLSCSSGFTISSYAMTWVRQAPGKGLFWAIIIMDSGDGH 60  
 DB 60 ADSAKGRFTISKDNANVTYLLQNSLKPEDTAVYYC-----NALITRWDKSYNDYWGQGT 114  
 QY 61 ADSVKGRTISRDNKNTLYLQNSLRPEDTGVYFCARDGSGHFCSSASCRGPDYWGQGT 120  
 DB 115 QVTSS 120  
 QY 121 PVTSS 126

## RESULT 7

HV3C HUMAN STANDARD; PRT; 119 AA.  
 ID HV3C HUMAN  
 AC P01764;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Ig heavy chain V-II region BUR.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 NC NCB1\_TaxID=9606;  
 RN [1]  
 RX MEDLINE=79151016; PubMed=107164;  
 RA Putnam F.W., Liu Y.-S.V., Low T.L.K.;  
 RT "Primary structure of a human IGA1 immunoglobulin. IV. Streptococcal  
 RT IGA1 protease, digestion, Fab and Fc fragments, and the complete  
 RT amino acid sequence of the alpha 1 heavy chain.";  
 RL J. Biol. Chem. 254:2865-2874(1979).  
 CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.  
 DR HSSP; P01772; 2F84.  
 DR GO; GO:0005576; C:extracellular; NAS.  
 DR GO; GO:0003823; P:antigen binding; NAS.  
 DR GO; GO:0006955; P:immune response; NAS.  
 DR InterPro; IPR007110; Ig-like.  
 DR InterPro; IPR003596; Ig\_V.  
 DR Pfam; PF00047; 1g\_1.  
 DR SMART; SMO0406; Igy; 1.  
 DR PROSITE; PS50835; IG\_LIKE; 1.  
 DR Immunoglobulin V region; Signal, 3D-structure.  
 FT SIGNAL 1 19  
 FT CHAIN 20 117 IG HEAVY CHAIN V-II REGION VH26.  
 FT DOMAIN 20 >117 IG-LIKE.  
 FT NON TER 117 117  
 SQ SEQUENCE 117 AA; 12582 MW; E826733F1A3CB9F1 CRC64;

Query Match 49.7%; Score 337.5; DB 1; Length 119;  
 Best Local Similarity 59.2%; Pred. No. 1.4e-25;  
 Matches 71; Conservative 14; Mismatches 32; Indels 3; Gaps 2;

QY 1 QVQLQESGGGLVQAGSLRLSCAASGTLISITIMDMYRQTPGKQRELYGRTITAGGS-TNY 59  
 DB 1 QVQLVSSGGGVQPGSLRLSCSSGFTISSYAMTWVRQAPGKGLFWAIIIMDSGDGH 60  
 QY 60 ADSAKGRFTISKDNANVTYLLQNSLKPEDTAVYYC-----NALITRWDKSYNDYWGQGT 119  
 DB 61 ADSVKGRTISRDNKNTLYLQNSLRPEDTGVYFCARDGSGHFCSSASCRGPDYWGQGT 118

## RESULT 8

HV3C HUMAN STANDARD; PRT; 117 AA.  
 ID HV3C HUMAN  
 AC P01764;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Ig heavy chain V-II region VH26 precursor.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 NC NCB1\_TaxID=9606;  
 RN [1]  
 RX MEDLINE=8101090; PubMed=6450418;  
 RA Mathysens G., Rabbits T.H.;  
 RT "Structure and multiplicity of genes for the human immunoglobulin  
 RT heavy chain variable region.";  
 RL Proc Natl. Acad. Sci. U.S.A. 77:6561-6565(1980).  
 CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; J00236; AA53516.1; -  
 DR EMBL; M35415; AA58735.1; -  
 DR PIR; A02047; H3H28.  
 DR PDB; 1HOU; 23-DEC-99.  
 DR Genew; HGNC:5545; IGHV@.  
 DR GO; GO:0005576; C:extracellular; NAS.  
 DR GO; GO:0003823; P:antigen binding; NAS.  
 DR GO; GO:0006955; P:immune response; NAS.  
 DR InterPro; IPR007110; Ig-like.  
 DR InterPro; IPR003596; Ig\_V.  
 DR Pfam; PF00047; 1g\_1.  
 DR SMART; SMO0406; Igy; 1.  
 DR PROSITE; PS50835; IG\_LIKE; 1.  
 DR Immunoglobulin V region; Signal, 3D-structure.  
 FT SIGNAL 1 19  
 FT CHAIN 20 117 IG HEAVY CHAIN V-II REGION VH26.  
 FT DOMAIN 20 >117 IG-LIKE.  
 FT NON TER 117 117  
 SQ SEQUENCE 117 AA; 12582 MW; E826733F1A3CB9F1 CRC64;

Query Match 49.3%; Score 334.5; DB 1; Length 117;  
 Best Local Similarity 70.8%; Pred. No. 2.6e-25;  
 Matches 68; Conservative 7; Mismatches 20; Indels 1; Gaps 1;

QY 1 QVQLQESGGGLVQAGSLRLSCAASGTLISITIMDMYRQTPGKQRELYGRTITAGGS-TNY 59  
 DB 20 EVQLVSSGGGVQPGSLRLSCSSGFTISSYAMTWVRQAPGKGLFWAIIIMDSGDGH 79  
 QY 60 ADSAKGRFTISKDNANVTYLLQNSLKPEDTAVYYC 95  
 DB 80 GDSVKGRTISRDNKNTLYLQNSLRPEDTAVYYC 115

RESULT 9  
 HV3C\_MOUSE

```

ID HV35 MOUSE STANDARD; PRT; 111 AA.
AC P01804;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ig heavy chain V-III region HPC76 (Fragment).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=8103937; PubMed=6251474;
RA Bernard O., Gough N.M.;
RT "Nucleotide sequence of immunoglobulin heavy chain joining segments
between translocated VH and mu constant regions genes."
RL Proc. Natl. Acad. Sci. U.S.A. 77:3630-3634(1980).
CC -1- MISCELLANEOUS: THE SEQUENCE OF THE FIRST 197 RESIDUES OF THE C
REGION WAS ALSO DETERMINED & DIFFERS IN ONLY 3 POSITIONS FROM THE
CORRESPONDING PORTION OF THE MOUSE MOPC 104E MU CHAIN.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.
DR HSSP; P01789; 1WCP.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003596; Ig_V.
DR Pfam; PF00047; Ig; 1.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS50835; IG-LIKE; 1.
KM Immunoglobulin V region.
FT DOMAIN 1
FT NON TER <1 110 IG-LIKE.
FT TER 111
SQ SEQUENCE 111 AA; 12304 MW; 0EDB98EC7348056A CRC64;

Query Match 49.1%; Score 334; DB 1; Length 111;
Best Local Similarity 58.1%; Pred. No. 3.2e-25;
Matches 68; Conservative 19; Mismatches 22; Indels 8; Gaps 3;

QY 6 EGGGGLVQAGSGSLRLSCAAGTILSIITMDWYRQTPGKRELVLG--RTAGGSTNYADA 63
DB 1 EGGGGLVQPGSGMKLSICVAGSFRTSNWMMWVROSPEKGLMEVAIRLKSGLATHYAESV 60
QY 64 KGRFTISDNKNTVYLOMNSLKPEDTAVYVCNALITRMKSVNDYNGCGTQTVSS 120
DB 61 KGRFTISDKSKSVYLOMNLRLADTGTTC---TR--PGVDYWGQGTTLTVSS 111

RESULT 10
HV02 CANPA STANDARD; PRT; 117 AA.
ID HV02 CANPA STANDARD; PRT; 117 AA.
AC P01785;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ig heavy chain V region MOO.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
NCBI_TaxID=9615;
RN [1]
RP SEQUENCE OF 1-112.
RX MEDLINE=77242268; PubMed=407924;
RA Wasserman R.L., Capra J.D.;
RT "Primary structure of the variable regions of two canine
immunoglobulin heavy chains."
RL Biochemistry 16:3160-3168(1977).
RN [2]
RP SEQUENCE OF 113-117.
RX MEDLINE=8007682; PubMed=117299;
RA McCumber L.J., Capra J.D.;
RT "The complete amino-acid sequence of a canine mu chain."
RL Mol. Immunol. 16:565-570(1979).
CC -1- MISCELLANEOUS: THIS MU CHAIN WAS ISOLATED FROM A MYELOMA PROTEIN.

```

```

CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.
DR PIR; A90403; MEDMO.
DR HSSP; P01772; 2F84.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003596; Ig_V.
DR Pfam; PF00047; Ig; 1.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS50835; IG-LIKE; 1.
KM Immunoglobulin V region.
FT DOMAIN 1
FT NON TER 117
FT TER 116 IG-LIKE.
SQ SEQUENCE 117 AA; 12703 MW; FCE07309E0A84B35 CRC64;

Query Match 49.1%; Score 333.5; DB 1; Length 117;
Best Local Similarity 58.3%; Pred. No. 3.2e-25;
Matches 70; Conservative 14; Mismatches 33; Indels 3; Gaps 1;

QY 1 QVQLQESGGGLVQAGSRLSCAAGTILSIITMDWYRQTPGKRELVRITAGGSTNYA 60
DB 1 EVLTLESQGGDLVPGGSLRLSCAAGSFRTSNMGMSVNRQDPEDGLQWVADISSGQTVYA 60
QY 61 DSAKGRFTISDNKNTVYLOMNSLKPEDTAVYVCNALITRMKSVNDYNGCGTQTVSS 120
DB 61 DAVKGRFTISDNKNTVYLOMNLRLVEDTAVYVC---ATGDIETIPRYGGQITLVSS 117

RESULT 11
HV39 MOUSE STANDARD; PRT; 118 AA.
ID HV39 MOUSE STANDARD; PRT; 118 AA.
AC P01809;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ig heavy chain V region X24.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxID=10090;
RN [1]
RP SEQUENCE.
RX MEDLINE=79223895; PubMed=111245;
RA Rao D.N., Rudikoff S., Kruttsch H., Potter M.;
RT "Structural evidence for independent joining region gene in
immunoglobulin heavy chains from anti-galactan myeloma proteins and
its potential role in generating diversity in
RT complementarity-determining regions."
RL Proc. Natl. Acad. Sci. U.S.A. 76:2890-2894(1979).
CC -1- MISCELLANEOUS: THIS CHAIN WAS ISOLATED FROM AN IGA MYELOMA PROTEIN
THAT BINDS GALACTAN.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.
DR PIR; F10394; AVMSX2.
DR HSSP; P01810; 2F8J.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003596; Ig_V.
DR Pfam; PF00047; Ig; 1.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS50835; IG-LIKE; 1.
KM Immunoglobulin V region.
FT DOMAIN 1
FT NON TER 118
FT TER 111 IG-LIKE.
SQ SEQUENCE 118 AA; 13105 MW; BB16A2DB677EF17F CRC64;

Query Match 49.1%; Score 333.5; DB 1; Length 118;
Best Local Similarity 59.0%; Pred. No. 3.2e-25;
Matches 72; Conservative 13; Mismatches 32; Indels 5; Gaps 3;

QY 1 QVQLQESGGGLVQAGSRLSCAAGTILSIITMDWYRQTPGKRELVRITAGGST-NY 59
DB 1 EVLTLESQGGDLVPGGSL-LSCAAGSFRTSNMGMSVNRQDPEDGLQWVADISSGQTVSS 59
QY 60 ADSAKGRFTISDNKNTVYLOMNSLKPEDTAVYVCNALITRMKSVNDYNGCGTQTVSS 119
DB 60 TPELKQKFTISDNKNTVYLOMNSKVRSEDTALVYCARL---GYGYPDYWGQGTTLTVS 116

```

QY 120 SE 121  
DB 117 SE 118

## RESULT 12

HV3U HUMAN STANDARD; PRT; 120 AA.  
AC P01782;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Ig heavy chain V-II region DOB.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE.  
RX MEDLINE=80020921; PubMed=114208;  
RA Steiner L.A., Garcia Pardo A., Margolies M.N.;  
RT "Amino acid sequence of the heavy-chain variable region of the  
RT crystallizable human myeloma protein Dob.";  
RT Biochemistry 18:4068-4080(1979).  
RN [2]  
RP CRYSTALLIZATION.  
RX MEDLINE=80020920; PubMed=114208;  
RA Steiner L.A., Lopes A.D.;  
RT "The crystallizable human myeloma protein Dob has a hinge-region  
RT deletion.";  
RT Biochemistry 18:4054-4067(1979).  
CC -1- MISCELLANEOUS: THIS GAMMA-1 MYELOMA PROTEIN HAS A DELETION IN THE  
CC HINGE REGION. THERE ARE NO LIGHT-HEAVY OR INTER-HEAVY CHAIN  
CC DISULFIDE BONDS.  
CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.  
DR PIR; A90431; GIHDB.  
DR HSSP; P01772; 2FB4.  
DR GO; GO:0005576; C:extracellular; NAS.  
DR GO; GO:0003823; F:antigen binding; NAS.  
DR GO; GO:0006955; P:immune response; NAS.  
DR InterPro; IPR007110; Ig-like.  
DR InterPro; IPR003596; Ig\_v.  
DR Pfam; PF00047; Ig\_1.  
DR SMART; SM00406; IGV; 1.  
DR PROSITE; PS50835; IG-LIKE; 1.  
KM Immunoglobulin V region.  
FT DOMAIN 1 112  
FT NON TER 120  
SQ SEQUENCE 120 AA; 13440 MW; 880DD307C432627 CRC64;

Query Match 49.0%; Score 333; DB 1; Length 120;  
Best Local Similarity 62.0%; Pred. No. 3,7e-25;  
Matches 75; Conservative 8; Mismatches 36; Indels 2; Gaps 2;

QY 1 QVOLOSGGGLVQAGSLRLSCAAGTILSIITMDWYRTPGKQRELVGRIT-AGGSTNY 59  
DB 1 EVQLVSGGDLVQPGKSLRLSCAAGFNRHEYNMELRQGPKGPEWSTITWNGSVLY 60  
QY 60 ADSAKRFITISKDNANTVYLQMSLKPEDTAVYVCNALITFMDSVNDYWGQGTQVTVS 119  
DB 61 ADSVKRFATISRDNAOKTLYLTQNLIRPEDTAVYVC-AKGIYINGNMTDSWGQGLTVTVS 119  
QY 120 S 120  
DB 120 S 120

RESULT 13  
HV01 CANFA STANDARD; PRT; 114 AA.  
ID HV01 CANFA  
AC P01784;  
DT 21-JUL-1986 (Rel. 01, Created)

DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Ig heavy chain V region GOM.  
OS Canis familiaris (Dog).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.  
NCBI\_TaxID=9615;  
RN [1]  
RP SEQUENCE.  
RX MEDLINE=77242268; PubMed=407924;  
RA Wasserman R.L., Capra J.D.;  
RT "Primary structure of the variable regions of two canine  
RT immunoglobulin heavy chains";  
RT Biochemistry 16:3160-3168(1977).  
CC -1- MISCELLANEOUS: THIS CHAIN WAS ISOLATED FROM A MYELOMA PROTEIN.  
CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.  
DR PIR; A02067; AVDGM.  
DR HSSP; P01772; 2FB4.  
DR InterPro; IPR007110; Ig-like.  
DR InterPro; IPR003596; Ig\_v.  
DR Pfam; PF00047; Ig\_1.  
DR SMART; SM00406; IGV; 1.  
DR PROSITE; PS50835; IG-LIKE; 1.  
KM Immunoglobulin V region.  
FT DOMAIN 1 112  
FT NON TER 114  
SQ SEQUENCE 114 AA; 12430 MW; B1D4745D2C4B13C4 CRC64;

Query Match 48.6%; Score 330; DB 1; Length 114;  
Best Local Similarity 59.5%; Pred. No. 6,8e-25;  
Matches 72; Conservative 10; Mismatches 31; Indels 8; Gaps 3;

QY 1 QVOLOSGGGLVQAGSLRLSCAAGTILSIITMDWYRTPGKQRELVGRIT-AGGSTNY 59  
DB 1 EVQLVSGGDLVQPGKSLRLSCAAGTITSGTDMQWRAPAPGKGLKAVYFNDALSAQY 60  
QY 60 ADSAKRFITISKDNANTVYLQMSLKPEDTAVYVCNALITFMDSVNDYWGQGTQVTVS 119  
DB 61 ADVKRFITISKDNADSLYLQMSLRADTAVYVC---APWQ---FHYWGQGLTVTVS 113  
QY 120 S 120  
DB 114 S 114

RESULT 14  
HV3B HUMAN STANDARD; PRT; 114 AA.  
ID HV3B HUMAN  
AC P01763;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Ig heavy chain V-II region WEA.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE.  
RX MEDLINE=83273707; PubMed=6410398;  
RA Goni F., Frangione B.;  
RT "Amino acid sequence of the Fv region of a human monoclonal IgM  
RT (protein WEA) with antibody activity against 3,4-pyruvylated  
RT galactose in Klebsiella polysaccharides K30 and K33.";  
RT Proc. Natl. Acad. Sci. U.S.A. 80:4837-4841(1983).  
CC -1- MISCELLANEOUS: THIS CHAIN WAS OBTAINED FROM A MONOCLONAL ANTIBODY  
CC AGAINST 3,4-PYRUVYLATED GALACTOSE AND ISOLATED FROM A PATIENT WITH  
CC WALDENSTROM'S MACROGLOBULINEMIA.  
CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.  
DR PIR; A02046; M3HWE.  
DR HSSP; P01772; 2FB4.  
DR GO; GO:0005576; C:extracellular; NAS.  
DR GO; GO:0003823; F:antigen binding; NAS.

DR GO:000695; P:immune response; NAS.  
 DR InterPro; IPR007110; Ig-like.  
 DR InterPro; IPR003596; Ig\_v.  
 DR Pfam; PF00047; Ig\_1.  
 DR SMART; SM00406; IGV; 1.  
 DR PROSITE; PS00835; IG LIKE; 1.  
 KW Immunoglobulin V region; Pyridone carboxylic acid.  
 FT DOMAIN 1 112  
 FT MOD RES 1 112 PYRROLIDONE CARBOXYLIC ACID.  
 FT NON TER 114 114  
 SQ SEQUENCE 114 AA; 12256 MW; D88294FB418A0787 CRC64;

Query Match 48.6%; Score 330; DB 1; Length 114;  
 Best Local Similarity 60.3%; Pred. No. 6.8e-25;  
 Matches 73; Conservative 14; Mismatches 26; Indels 8; Gaps 4;

QY 1 QVQLQESGGGLVQAGSRLRSCASGTTLSITIMDWRQTGPKORELYGRTTAGST-NY 59  
 DB 1 QVQLVDSGGGLVPEPGSLRLSCASGFTFSANDMNMWRQAPGKLEWLSFIGSGSTIYY 60  
 QY ADSAKGRFTISKDNKNTVYLQWNSLKPEDTAVYYCNALITRWDKSVNDYWGOGTQVTVS 119  
 DB 61 ADSVKGRTTISKDNKNTVYLQWNSLKPEDTAVYYC--ARGW-LIN--WGQGLTVTVS 113  
 QY 120 S 120  
 DB 114 S 114

## RESULT 15

HV37\_MOUSE STANDARD; PRT; 119 AA.  
 ID HV37\_MOUSE P01807;  
 AC P01807;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Ig heavy chain V region X44.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 OX NCBI\_TaxId=10090;  
 RN [1]  
 RP SEQUENCE.  
 RX MEDLINE=79223895; Pubmed=11245;  
 RA Rao D.N., Rudikoff S., Krutzsch H., Potter M.;  
 RT "Structural evidence for independent joining region gene in  
 RT immunoglobulin heavy chains from anti-galactan myeloma proteins and  
 RT its potential role in generating diversity in  
 RT complementarity-determining regions";  
 RL Proc. Natl. Acad. Sci. U.S.A. 76:2890-2894(1979).  
 CC -1- MISCELLANEOUS: THIS CHAIN WAS ISOLATED FROM AN IGA MYELOMA PROTEIN  
 CC THAT BINDS GALACTAN.  
 CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.  
 DR PIR; A02077; AVMSX4.  
 DR HSSP; P018-0; 2FBJ.  
 DR InterPro; IPR007110; Ig-like.  
 DR InterPro; IPR003596; Ig\_v.  
 DR Pfam; PF00047; Ig\_1.  
 DR SMART; SM00406; IGV; 1.  
 DR PROSITE; PS00835; IG LIKE; 1.  
 KW Immunoglobulin V region.  
 FT DOMAIN 1 117  
 FT NON TER 119 119 IG-LIKE.  
 SQ SEQUENCE 119 AA; 13246 MW; BC34FC6F31CD41B3 CRC64;

Query Match 48.6%; Score 330; DB 1; Length 119;  
 Best Local Similarity 57.4%; Pred. No. 7.1e-25;  
 Matches 70; Conservative 16; Mismatches 32; Indels 4; Gaps 3;

QY 1 QVQLQESGGGLVQAGSRLRSCASGTTLSITIMDWRQTGPKORELYGRTTAGST-NY 59  
 DB 1 EVKLLESGGGLVPEPGSLRLSCASGFTFSRYNMWVRQAPGKLEWISINPDSTIYY 60

QY 60 ADSAKGRFTISKDNKNTVYLQWNSLKPEDTAVYYCNALITRWDKSVNDYWGOGTQVTVS 119  
 DB 61 TPLSKDKFTISKDNKNTVYLQWNSLKVEDTALYYC-ASLHYGYAA--YWGQGLTVTVS 117  
 QY 120 SE 121  
 DB 118 AE 119

Search completed: October 5, 2004, 08:01:54  
 Job time : 11.5463 secs

The Page Order (up to)





Db 1 EVOLVSGGGGLVOPGSLRLSCAASGFTYSNNMVRQAPGKGLSESVITVSSGSSYYA 60  
 QY 61 DSAKGRFTISKDNKNTVYLQNNSLKPEPTAVYYCNALITRMDKSVNDVWGCTQVTVSS 120  
 61 DSAKGRFTISKDNKNTVYLQNNSLKPEPTAVYYCAR--DREGELFDVWGCTLVTVSS 118

## RESULT 2

Q96BB9 PRELIMINARY; PRT; 597 AA.  
 ID Q96BB9  
 AC Q96BB9  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, last annotation update)  
 DE Hypothetical protein.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=B-cell;  
 RA Strausberg R.;  
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC015760; AA015760.1; -.  
 DR InterPro; IPR007110; IG\_1like.  
 DR InterPro; IPR003006; IG\_MHC.  
 DR InterPro; IPR003596; IG\_V.  
 DR Pfam; PF00047; IG\_5.  
 DR SMART; SM00406; IGV\_1.  
 DR PROSITE; PS00835; IG\_LIKE; 5.  
 DR PROSITE; PS00290; IG\_MHC; 3.  
 KM Hypothetical protein.  
 SQ SEQUENCE 597 AA; 65039 MW; 4FCA3AD8CE263D9 CRC64;

Query Match 55.7%; Score 378.5; DB 4; Length 597;  
 Best Local Similarity 64.8%; Pred. No. 2.4e-30;  
 Matches 81; Conservative 11; Mismatches 28; Indels 5; Gaps 2;

QY 1 QVLOESGGGLVOPGSLRLSCAASGFTISIIYMDVYRQTPGKQRELVGRI--TAGSGT 59  
 20 EVOLVSGGGGLVOPGSLRLSCAASGFTYSNNMVRQAPGKGLSESVITVSSGSSYY 79  
 Db 60 DSAKGRFTISKDNKNTVYLQNNSLKPEPTAVYYCNALITRMDKSVN---DYGCGGQ 115  
 80 DSAKGRFTISKDNKNTVYLQNNSLKPEPTAVYYCAR--DREGELFDVWGCTLVTVSS 139  
 QY 116 VTVSS 120  
 Db 140 VTVSS 144

## RESULT 3

Q9ULB8 PRELIMINARY; PRT; 131 AA.  
 ID Q9ULB8  
 AC Q9ULB8  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, last annotation update)  
 DE Myosin-reactive immunoglobulin heavy chain variable region (Fragment).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=98271139; PubMed=9614934;  
 RA Wu X., Liu B., Van der Merwe P.L., Kalis N.N., Berney S.M.,  
 RA Young D.C.;  
 RT "Myosin-reactive autoantibodies in rheumatic carditis and normal  
 RT fetus";  
 RL Clin. Immunol. Immunopathol. 87:184-192 (1998).

DR EMBL; AF035026; AAD56262.1; -.  
 DR PIR; S21205; S21205.  
 DR HSPD; P01810; 2REJ.  
 DR InterPro; IPR007110; IG\_1like.  
 DR InterPro; IPR003596; IG\_V.  
 DR Pfam; PF00047; IG\_1.  
 DR SMART; SM00406; IGV\_1.  
 DR PROSITE; PS00835; IG\_LIKE; 1.  
 FT NON\_TER 1  
 FT NON\_TER 131  
 SQ SEQUENCE 131 AA; 14142 MW; 96E7D668E375DEA0 CRC64;

Query Match 54.3%; Score 372.5; DB 4; Length 131;  
 Best Local Similarity 61.1%; Pred. No. 1.3e-30;  
 Matches 80; Conservative 13; Mismatches 27; Indels 11; Gaps 3;

QY 1 QVLOESGGGLVOPGSLRLSCAASGFTISIIYMDVYRQTPGKQRELVGRI--TAGSGT 57  
 1 EVOLVSGGGGLVOPGSLRLSCAASGFTISKAMNSVRQAPGKGLSESVITVSSGSSYY 60  
 Db 58 NYADSAKGRFTISKDNKNTVYLQNNSLKPEPTAVYYCN-----ALITRMDKSVN-DY 109  
 61 DYAPVKGRLITRDSKNTLYLRNLSLKTEDTAVYYCTGTIMIVITTSKRTSPFY 120  
 QY 110 WGCTQVTVSS 120  
 Db 121 WGCTLVTVSS 131

## RESULT 4

Q91XEL PRELIMINARY; PRT; 480 AA.  
 ID Q91XEL  
 AC Q91XEL  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, last annotation update)  
 DE Hypothetical protein (Fragment).  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Colon;  
 RA Strausberg R.;  
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC010798; AA010798.1; -.  
 DR InterPro; IPR007110; IG\_1like.  
 DR InterPro; IPR003006; IG\_MHC.  
 DR InterPro; IPR003596; IG\_V.  
 DR Pfam; PF00047; IG\_4.  
 DR SMART; SM00406; IGV\_1.  
 DR PROSITE; PS00835; IG\_LIKE; 4.  
 DR PROSITE; PS00290; IG\_MHC; 2.  
 KM Hypothetical protein.  
 FT NON\_TER 1  
 FT NON\_TER 480  
 SQ SEQUENCE 480 AA; 51936 MW; 20B9234BEF2B41ED CRC64;

Query Match 54.3%; Score 368.5; DB 11; Length 480;  
 Best Local Similarity 65.6%; Pred. No. 1.9e-29;  
 Matches 80; Conservative 9; Mismatches 28; Indels 5; Gaps 2;

QY 2 VLOESGGGLVOPGSLRLSCAASGFTISIIYMDVYRQTPGKQRELVGRI--TAGSGT 60  
 20 VLOESGGGLVOPGSLRLSCAASGFTISIIYMDVYRQTPGKQRELVGRI--TAGSGT 79  
 Db 61 DSAKGRFTISKDNKNTVYLQNNSLKPEPTAVYYCNALITRMDKSVNDVWGCTQVTVSS 120  
 80 DSAKGRFTISKDNKNTVYLQNNSLKPEPTAVYYCAR--DREGELFDVWGCTLVTVSS 135  
 QY 121 EP 122  
 Db 136 EP 137

RESULT 5

Q8WTK1 PRELIMINARY: PRT; 613 AA.

ID Q8WTK1

AC Q8WTK1

DT 01-MAR-2002 (TrEMBLrel. 20, Created)

DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Hypothetical protein.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

NCBI\_TaxID=9606;

Query Match

Best local Similarity 53.3%; Score 362; DB 4; Length 613;

Matches 79; Conservative 7; Mismatches 31; Indels 6; Gaps 3;

QY 1 QVQLQESGGGLVQAGSRLRSCAASGTLISITIMDWRQTPGKQRELVGRITAGSSTN-Y 59

DB 20 QVQLVESGGGVQVQPSRLRLSCAASGTFSSYGMHWRAQPKGLEWAVSYGSKYY 79

QY 60 ADSAKGRFTISKDNANTVYLQNSLKPEDTAVYVCNALITRMDKSVN--DYWGCGTQVT 117

DB 80 ADVSKGRFTISKDNANTVYLQNSLKPEDTAVYVCNALITRMDKSVN--DYWGCGTQVT 117

QY 118 VSS 120

DB 137 VSS 139

RESULT 6

Q9UL90 PRELIMINARY: PRT; 113 AA.

ID Q9UL90

AC Q9UL90

DT 01-MAY-2000 (TrEMBLrel. 13, Created)

DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Myosin-reactive immunoglobulin heavy chain variable region (Fragment).

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

NCBI\_TaxID=9606;

QY 1 QVQLQESGGGLVQAGSRLRSCAASGTLISITIMDWRQTPGKQRELVGRITAGSSTN-Y 59

DB 1 EVQLVESGGGVQVQPSRLRLSCAASGTFSSYGMHWRAQPKGLEWAVSYGSKYY 79

QY 60 ADSAKGRFTISKDNANTVYLQNSLKPEDTAVYVCNALITRMDKSVN--DYWGCGTQVT 117

DB 80 ADVSKGRFTISKDNANTVYLQNSLKPEDTAVYVCNALITRMDKSVN--DYWGCGTQVT 117

QY 118 VSS 120

DB 137 VSS 139

DR InterPro; IPR003596; Ig\_V.

DR Pfam; PF00047; Ig; 1.

DR SMART; SM00406; IgV; 1.

DR PROSITE; PSS0835; IG\_LIKE; 1.

FT NON\_TER 1

FT NON\_TER 113

QY 1 QVQLQESGGGLVQAGSRLRSCAASGTLISITIMDWRQTPGKQRELVGRITAGSSTN-Y 59

DB 1 EVQLVESGGGVQVQPSRLRLSCAASGTFSSYGMHWRAQPKGLEWAVSYGSKYY 79

QY 60 ADSAKGRFTISKDNANTVYLQNSLKPEDTAVYVCNALITRMDKSVN--DYWGCGTQVT 117

DB 61 ADVSKGRFTISKDNANTVYLQNSLKPEDTAVYVCNALITRMDKSVN--DYWGCGTQVT 117

QY 120 S 120

DB 113 S 113

RESULT 7

Q9UL71 PRELIMINARY: PRT; 121 AA.

ID Q9UL71

AC Q9UL71

DT 01-MAY-2000 (TrEMBLrel. 13, Created)

DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Myosin-reactive immunoglobulin heavy chain variable region (Fragment).

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

NCBI\_TaxID=9606;

QY 1 QVQLQESGGGLVQAGSRLRSCAASGTLISITIMDWRQTPGKQRELVGRITAGSSTN-Y 59

DB 1 EVQLVESGGGVQVQPSRLRLSCAASGTFSSYGMHWRAQPKGLEWAVSYGSKYY 79

QY 60 ADSAKGRFTISKDNANTVYLQNSLKPEDTAVYVCNALITRMDKSVN--DYWGCGTQVT 117

DB 80 ADVSKGRFTISKDNANTVYLQNSLKPEDTAVYVCNALITRMDKSVN--DYWGCGTQVT 117

QY 118 VSS 120

DB 137 VSS 139

Query Match

Best local Similarity 52.9%; Score 359.5; DB 4; Length 121;

Matches 77; Conservative 12; Mismatches 29; Indels 5; Gaps 3;

QY 1 QVQLQESGGGLVQAGSRLRSCAASGTLISITIMDWRQTPGKQRELVGRITAGSSTN-Y 59

DB 1 EVQLVESGGGVQVQPSRLRLSCAASGTFSSYGMHWRAQPKGLEWAVSYGSKYY 79

QY 60 ADSAKGRFTISKDNANTVYLQNSLKPEDTAVYVCNALITRMDKSVN--DYWGCGTQVT 117

DB 61 ADVSKGRFTISKDNANTVYLQNSLKPEDTAVYVCNALITRMDKSVN--DYWGCGTQVT 117

QY 118 VSS 120

DB 137 VSS 139

```

RESULT 8
ID Q9NSK4 PRELIMINARY; PRT; 499 AA.
AC Q9NSK4;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Blood;
RA Strausberg R.;
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC032249; AAH32249.1; -.
DR InterPro: IPR003599; Ig_1.
DR InterPro: IPR007110; Ig_1like.
DR InterPro: IPR003597; Ig_G1.
DR InterPro: IPR003006; Ig_MHC.
DR InterPro: IPR003596; Ig_V.
DR Pfam: PF00047; Ig_4.
DR SMART: SM00409; IG1.4.
DR SMART: SM00407; IG1.2.
DR PROSITE: PS00835; IG_LIKE; 4.
DR PROSITE: PS00290; IG_MHC; 1.
KM Hypothetical protein.
SQ SEQUENCE 499 AA; 53376 MW; 93A5C89582054F32 CRC64;

Query Match 52.9%; Score 359; DB 4; Length 499;
Best Local Similarity 57.7%; Pred. No. 2e-28;
Matches 79; Conservative 13; Mismatches 37; Indels 8; Gaps 3;

QY 1 QVQLQESGGGLVQAGSLRLSCAASGTTISIYYMDVYRQTPGKQRELVGRIT-AGSGTNY 59
DB 20 EVQLVESGGGVVPPGSLRLSCATSGFTPDGSAVWRQAPGKGLWVSISIMNGSGSTNY 79
QY 60 ADSAKGRFTISKNAKNTVYLQWNSLKPEDTAVYVCNALITRMDKSVN-----DYWGQG 113
DB 80 ADVKGRFTISRDNANKSYLQWNSLRVEDTALVYCARPTPKYCGSGSLGYMDVWKGK 139
QY 114 TQYTVASS-EKTPKPOP 129
DB 140 TTVTVSSASPTSPKVP 156

RESULT 9
ID Q9UL91 PRELIMINARY; PRT; 118 AA.
AC Q9UL91;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Myosin-reactive immunoglobulin heavy chain variable region
DE (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=88277139; PubMed=9614934;
RA Wu X., Liu B., Van der Merwe P.L., Kalis N.N., Berner S.M.,
RA Young D.C.;
RT "Myosin-reactive autoantibodies in rheumatic carditis and normal
RT fetus.";
RL Clin. Immunol. Immunopathol. 87:184-192(1998).
DR EMBL: AF035023; AAD56259.1; -.
DR PIR: S21205; S21205.

```

```

DR HSSP: P01772; 2FB4.
DR InterPro: IPR007110; Ig_1like.
DR InterPro: IPR003596; Ig_V.
DR Pfam: PF00047; Ig_1.
DR SMART: SM00406; IG1.1.
DR PROSITE: PS00835; IG_LIKE; 1.
FT NON_TER 1
FT NON_TER 118
SQ SEQUENCE 118 AA; 12643 MW; D0633949F2AC149D CRC64;

Query Match 52.4%; Score 356; DB 4; Length 118;
Best Local Similarity 65.3%; Pred. No. 5.9e-29;
Matches 79; Conservative 8; Mismatches 28; Indels 6; Gaps 3;

QY 1 QVQLQESGGGLVQAGSLRLSCAASGTTISIYYMDVYRQTPGKQRELVGRITAGSGT-NY 59
DB 1 EVQLVESGGGLVPPGSLRLSCAASGFTSSVMMVWRQAPGKGLWVSISITITTY 60
QY 60 ADSAKGRFTISKNAKNTVYLQWNSLKPEDTAVYVCNALITRMDKSVN-VNDYWGQGYQTV 118
DB 61 ADVKGRFTISRDNANKSYLQWNSLRVEDTALVYCARPTPKYCGSGSLGYMDVWKGK 116
QY 119 S 119
DB 117 S 117

RESULT 10
ID Q920E7 PRELIMINARY; PRT; 119 AA.
AC Q920E7;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Pterin-mimicking anti-idiotypic heavy chain variable region
DE (Fragment).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Atkin J.D., Iape A., Jennings I.G., Horatits O., Cotton R.G.H.;
RT "Definition of the Idiotypic of Pterin-mimicking Antibodies Expressed
RT in Mammalian Cells.";
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF307937; AAL09421.1; -.
DR PIR: C25913; C25913.
DR InterPro: IPR007110; Ig_1like.
DR InterPro: IPR003596; Ig_V.
DR Pfam: PF00047; Ig_1.
DR SMART: SM00406; IG1.1.
DR PROSITE: PS00835; IG_LIKE; 1.
FT NON_TER 1
FT NON_TER 119
SQ SEQUENCE 119 AA; 13025 MW; F6E90404381CA7C CRC64;

Query Match 52.1%; Score 353.5; DB 11; Length 119;
Best Local Similarity 63.1%; Pred. No. 1.1e-28;
Matches 77; Conservative 11; Mismatches 29; Indels 5; Gaps 3;

QY 1 QVQLQESGGGLVQAGSLRLSCAASGTTISIYYMDVYRQTPGKQRELVGRITAGGS-TNY 59
DB 1 EVQLVESGGGLVPPGSLRLSCAASGFTSSVMMVWRQAPGKGLWVSISITITTY 60
QY 60 ADSAKGRFTISKNAKNTVYLQWNSLKPEDTAVYVCNALITRMDKSVN-DYWGQGYQTV 118
DB 61 PDSVKGKFTISRDNANKSYLQWNSLRVEDTALVYCARPTPKYCGSGSLGYMDVWKGK 117
QY 119 SS 120
DB 118 SA 119

```

```

RESULT 11
ID Q99KA4 PRELIMINARY; PRT; 487 AA.
AC Q99KA4
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DE 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC004786; AA04786.1; -.
DR HSSP; P01810; 2FRT.
DR InterPro; IPR007110; IG_1like.
DR InterPro; IPR003006; IG_MHC.
DR InterPro; IPR003596; IG_V.
DR Pfam; PF00047; IG; 4.
DR SMART; SMO0406; IGV; 1.
DR PROSITE; PSS0835; IG_LIKE; 4.
DR PROSITE; PSS0290; IG_MHC; 2.
KM Hypothetical protein.
SQ SEQUENCE 487 AA; 52554 MW; 7DC8B36DB33077B CRC64;

Query Match
Best Local Similarity 52.0%; Score 353; DB 11; Length 487;
Matches 76; Conservative 14; Mismatches 26; Indels 14; Gaps 3;

QY 1 QVQLQESGGGLVQAGSLRLSCAASGTLISITIMDVRQTPGKQRELVGRITAGGS-TNY 59
DB 20 EVQLVESGGGLVQPGGSLRLSCAASGTFSSYAMSWVRQATEKRLIEWALTISDGSITYY 79
QY 60 ADSAKGRFTISKDNKNTVYLQMSLKPEDTAVYYCNALITPMDSVNDYWG 111
DB 80 PDNVKGRFTISRDNKNTLFLQMTSLRSDPTAMYYCAIDMGSPYGSISRF-----DYWG 134
QY 112 QGTQVTVSSE 121
DB 135 QGTITVSSSE 144

RESULT 12
ID Q91205 PRELIMINARY; PRT; 473 AA.
AC Q91205;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DE 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC010327; AA010327.1; -.
DR MGD; MGI:2144967; A0044919.
DR GO; GO:0005489; F:electron transporter activity; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro; IPR00345; Cytochrome_B5.
DR InterPro; IPR007110; IG_1like.
DR InterPro; IPR003006; IG_MHC.
DR InterPro; IPR003596; IG_V.
DR Pfam; PF00047; IGV; 3.
DR SMART; SMO0406; IGV; 1.

```

```

DR PROSITE; PSS0190; CYTOCHROME_C; 1.
DR PROSITE; PSS0835; IG_LIKE; 4.
DR PROSITE; PSS0290; IG_MHC; 1.
KM Hypothetical protein.
SQ SEQUENCE 473 AA; 51946 MW; CFE625F008932A2A12 CRC64;

Query Match
Best Local Similarity 51.5%; Score 350; DB 11; Length 473;
Matches 74; Conservative 15; Mismatches 33; Indels 4; Gaps 3;

QY 1 QVQLQESGGGLVQAGSLRLSCAASGTLISITIMDVRQTPGKQRELVGRITAGGS-TNY 59
DB 20 EVQLVESGGGLVQPGGSLRLSCAASGTFSDYGMHWVRQAPKGLVWVAYINSGSTITYY 79
QY 60 ADSAKGRFTISKDNKNTVYLQMSLKPEDTAVYYCNALITPMDSVNDYWGQGTQVTVS 119
DB 80 ADIVKGRFTISRDNKNTLFLQMTSLRSDPTAMYYCAREL--WLRII-DYWGQGITITVS 136
QY 120 SEPKTP 125
DB 137 SAKTTP 142

RESULT 13
ID Q91207 PRELIMINARY; PRT; 486 AA.
AC Q91207;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DE 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC010324; AA010324.1; -.
DR InterPro; IPR007110; IG_1like.
DR InterPro; IPR003006; IG_MHC.
DR InterPro; IPR003596; IG_V.
DR Pfam; PF00047; IG; 4.
DR SMART; SMO0406; IGV; 1.
DR PROSITE; PSS0835; IG_LIKE; 4.
DR PROSITE; PSS0290; IG_MHC; 2.
KM Hypothetical protein.
SQ SEQUENCE 486 AA; 52682 MW; 4FEF835125DA870B CRC64;

Query Match
Best Local Similarity 51.3%; Score 348.5; DB 11; Length 486;
Matches 71; Conservative 16; Mismatches 34; Indels 3; Gaps 1;

QY 1 QVQLQESGGGLVQAGSLRLSCAASGTLISITIMDVRQTPGKQRELVGRITAGGSTNYA 60
DB 20 EVQLVESGGGLVQPGGSLRLSCVSGFSTYDMSWVRQTEPRLEWVAITSGNTIYP 79
QY 61 DSAKGRFTISKDNKNTVYLQMSLKPEDTAVYYC--NALITRMKSVNDYWGQGTQV 117
DB 80 DNVKGRFTISRDNKNTLFLQMTSLRSDPTAMYYCAVEPIPIYYSGSFDSWGQGITIT 139
QY 118 VSSE 121
DB 140 VSSE 143

RESULT 14
ID Q9HCCI PRELIMINARY; PRT; 112 AA.
AC Q9HCCI;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)

```

DT 01-OCT-2003 (T-EMBLrel. 25, last annotation update)  
 DE Single chain Fv (Fragment).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Kikuchi M., Takeda C., Tsujimoto Y., Asada S., Nagata K.,  
 RT "An antibody fragment2A3 specific for native lysozyme: Isolation from a  
 human synthetic phage display library and characterization."  
 RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AB049915; BAB16829.1; -  
 DR HSSP; P01772; 2FB4.  
 DR InterPro; IPR007110; Ig-like.  
 DR InterPro; IPR003596; Ig\_V.  
 DR Pfam; PF00047; Ig\_1.  
 DR SMART; SM00406; IGV\_1.  
 DR PROSITE; PS50835; IG\_LIKE; 1.  
 FT NON\_TER 1  
 FT TER 112  
 SQ SEQUENCE 112 AA; 12243 MW; 24F1A45EC3B84788 CRC64;  
 Query Match 51.1%; Score 347; DB 4; Length 112;  
 Best local Similarity 62.4%; Pred. No. 4.6e-28;  
 Matches 73; Conservative 9; Mismatches 29; Indels 6; Gaps 2;  
 QY 1 QVLOLSSGGSLVAGSGSLRSCASGTLISIIYMDVRYOTPGKRELVGRII-AGGSTNY 59  
 DB 1 EVOLVSSGGSLVAPGSLRISCSAGFTPDYGMKVRQAPKGLFWVSGIINNGSTGY 60  
 QY 60 ADSAKGRFTISKDNKNTVYLQNSLKPEDTAVVYC-----NALITRMDSVNDYWGQGT 114  
 DB 61 ADSVKGKRFITSRDNKNSLYLQNSLRADTAVVYCAR-----RYALDYWGQGTIV 112  
 RESULT 15  
 Q8WU38 PRELIMINARY; PRT; 573 AA.  
 AC Q8WU38;  
 DT 01-MAR-2002 (T-EMBLrel. 20, Created)  
 DT 01-MAR-2002 (T-EMBLrel. 20, Last sequence update)  
 DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)  
 DE Hypothetical protein.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA TISSUE=tonsil;  
 RA Strausberg R.;  
 RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC021276; AAH21276.1; -  
 DR PIR; S21205;  
 DR InterPro; IPR007110; Ig-like.  
 DR InterPro; IPR003006; IG\_MHC.  
 DR InterPro; IPR003596; IG\_V.  
 DR Pfam; PF00047; Ig\_4.  
 DR SMART; SM00406; IGV\_1.  
 DR PROSITE; PS50835; IG\_LIKE; 4.  
 DR PROSITE; PS00290; IG\_MHC; 2.  
 KW Hypothetical protein.  
 SQ SEQUENCE 573 AA; 62967 MW; FD072344033AC530 CRC64;  
 Query Match 50.9%; Score 345.5; DB 4; Length 573;  
 Best local Similarity 58.8%; Pred. No. 5.9e-27;  
 Matches 80; Conservative 9; Mismatches 38; Indels 9; Gaps 4;  
 QY 1 QVLOLSSGGSLVAGSGSLRSCASGTLISIIYMDVRYOTPGKRELVGRII-AGGSTNY 59  
 DB 20 EVOLVSSGGSLVAPGSLRISCSAGFTPDYGMKVRQAPKGLFWVSGIINNGSTGY 79

QY 60 ADSAKGRFTISKDNKNTVYLQNSLKPEDTAVVYC-----NALITRMDSVNDYWGQGT 114  
 DB 80 ADSVKGKRFITSRDNKNSLYLQNSLRADTAVVYCAR-----RYALDYWGQGT 137  
 QY 115 QVTVSSSEP-KTPKPOP 129  
 DB 138 TVTVSSAPTKAPDVFP 153

Search completed: October 5, 2004, 08:13:45  
 Job time : 69.2824 secs



XX The patent discloses antibodies or their fragments comprising a heavy  
 CC chain variable domain (VH) derived from an immunoglobulin naturally  
 CC devoid of light chains specific for inhibiting human dietary enzymes. The  
 CC antibodies of the invention are useful for the preparation of medicaments  
 CC or food for inhibiting the activity of one or more human dietary enzymes  
 CC especially human pancreatic lipase (HPL) or human gastric lipase (HGL)  
 CC which are useful for the cosmetic control of body weight of human beings.  
 CC The present peptide sequence is HPL inhibiting VH fragment, HPL #13 from  
 CC llama (camelid) species  
 XX  
 SQ Sequence 129 AA;  
 Query Match 100.0%; Score 679; DB 4; Length 129;  
 Best Local Similarity 100.0%; Pred. No. 4,6e-52;  
 Matches 129; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 QVQLQESGGGLVQAGSLRLSCAASGTLISITIMDYRQTPGKRELVGRTTAGSTNYA 60  
 DB 1 QVQLQESGGGLVQAGSLRLSCAASGTLISITIMDYRQTPGKRELVGRTTAGSTNYA 60  
 QY 61 DSAKGRFTISKDNKNTVYLQNNSLKPEPTAVYYCNALITRMDKSVNDYWGQGTQVTVSS 120  
 DB 61 DSAKGRFTISKDNKNTVYLQNNSLKPEPTAVYYCNALITRMDKSVNDYWGQGTQVTVSS 120  
 QY 121 EPKTPKPOP 129  
 DB 121 EPKTPKPOP 129  
 RESULT 2  
 ID AAE10558 standard; peptide; 129 AA.  
 AC AAE10558;  
 XX  
 DT 10-DEC-2001 (first entry)  
 XX  
 DE HPL inhibiting VH fragment, HPL #22 from llama species.  
 XX  
 KM llama antibody; camelid; anorectic; heavy chain variable domain; VH;  
 KM human dietary enzyme inhibitor; medicament; human pancreatic lipase; HPL;  
 KM food; human gastric lipase; HGL; cosmetic control; body weight.  
 XX  
 OS llama sp.  
 XX  
 FH Location/Qualifiers  
 FT 31..35  
 FT Region /label=CDR1  
 FT /note="Complementarity determining region 1"  
 FT 50..64  
 FT Region /label=CDR2  
 FT /note="Complementarity determining region 2"  
 FT 98..109  
 FT Region /label=CDR3  
 FT /note="Complementarity determining region 3"  
 XX  
 PN EPI134231-A1.  
 XX  
 PD 19-SEP-2001.  
 XX  
 PF 20-FEB-2001; 2001EP-00200703.  
 XX  
 PR 14-MAR-2000; 2000EP-00200930.  
 XX  
 PA (UNITL) UNILEVER NV.  
 PA (UNITL) UNILEVER PLC.  
 XX  
 PI Bezemer S, Van De Burg M, De Haard JW, Tarellus E;  
 XX  
 DR WPI, 2001-572718/65.  
 XX  
 PT New antibody or its fragments for inhibiting human dietary enzymes,

PT useful for cosmetic control of body weight of human beings, comprises  
 FT heavy chain variable domain derived from immunoglobulin naturally devoid  
 FT of light chains.  
 XX  
 PS Example 2; Page 10; 37pp; English.  
 XX  
 CC The patent discloses antibodies or their fragments comprising a heavy  
 CC chain variable domain (VH) derived from an immunoglobulin naturally  
 CC devoid of light chains specific for inhibiting human dietary enzymes. The  
 CC antibodies of the invention are useful for the preparation of medicaments  
 CC or food for inhibiting the activity of one or more human dietary enzymes  
 CC especially human pancreatic lipase (HPL) or human gastric lipase (HGL)  
 CC which are useful for the cosmetic control of body weight of human beings.  
 CC The present peptide sequence is HPL inhibiting VH fragment, HPL #22 from  
 CC llama (camelid) species  
 XX  
 SQ Sequence 129 AA;  
 Query Match 79.1%; Score 537; DB 4; Length 129;  
 Best Local Similarity 79.8%; Pred. No. 1,4e-39;  
 Matches 103; Conservative 11; Mismatches 15; Indels 0; Gaps 0;  
 QY 1 QVQLQESGGGLVQAGSLRLSCAASGTLISITIMDYRQTPGKRELVGRTTAGSTNYA 60  
 DB 1 QVQLQESGGGLVQAGSLRLSCAASGTLISITIMDYRQTPGKRELVGRTTAGSTNYA 60  
 QY 61 DSAKGRFTISKDNKNTVYLQNNSLKPEPTAVYYCNALITRMDKSVNDYWGQGTQVTVSS 120  
 DB 61 DSAKGRFTISKDNKNTVYLQNNSLKPEPTAVYYCNALITRMDKSVNDYWGQGTQVTVSS 120  
 QY 121 EPKTPKPOP 129  
 DB 121 EPKTPKPOP 129  
 RESULT 3  
 ID AAE10555 standard; peptide; 131 AA.  
 AC AAE10555;  
 XX  
 DT 10-DEC-2001 (first entry)  
 XX  
 DE HPL inhibiting VH fragment, HPL #15 from llama species.  
 XX  
 KM llama antibody; camelid; anorectic; heavy chain variable domain; VH;  
 KM human dietary enzyme inhibitor; medicament; human pancreatic lipase; HPL;  
 KM food; human gastric lipase; HGL; cosmetic control; body weight.  
 XX  
 OS llama sp.  
 XX  
 FH Location/Qualifiers  
 FT 31..35  
 FT Region /label=CDR1  
 FT /note="Complementarity determining region 1"  
 FT 50..64  
 FT Region /label=CDR2  
 FT /note="Complementarity determining region 2"  
 FT 98..111  
 FT Region /label=CDR3  
 FT /note="Complementarity determining region 3"  
 XX  
 PN EPI134231-A1.  
 XX  
 PD 19-SEP-2001.  
 XX  
 PF 20-FEB-2001; 2001EP-00200703.  
 XX  
 PR 14-MAR-2000; 2000EP-00200930.  
 XX  
 PA (UNITL) UNILEVER NV.  
 PA (UNITL) UNILEVER PLC.  
 XX



PI Bezemer S, Van De Burg M, De Haard JWW, Tareilus E;  
XX WPI; 2001-572718/65.  
XX  
XX  
XX New antibody or its fragments for inhibiting human dietary enzymes,  
PT useful for cosmetic control of body weight of human beings, comprises  
PT heavy chain variable domain derived from immunoglobulin naturally devoid  
PT of light chains.  
XX  
XX Example 2; Page 9; 37pp; English.  
XX  
XX The patent discloses antibodies or their fragments comprising a heavy  
CC chain variable domain (VH) derived from an immunoglobulin naturally  
CC devoid of light chains specific for inhibiting human dietary enzymes. The  
CC antibodies of the invention are useful for the preparation of medicaments  
CC or food for inhibiting the activity of one or more human dietary enzymes  
CC especially human pancreatic lipase (HPL) or human gastric lipase (HGL)  
CC which are useful for the cosmetic control of body weight of human beings.  
CC The present peptide sequence is HPL inhibiting VH fragment, HPL #15 from  
CC llama (camelid) species  
XX  
XX Sequence 131 AA;  
SQ  
Query Match 78.4%; Score 532; DB 4; Length 131;  
Best Local Similarity 80.9%; Pred. No. 3.9e-39;  
Matches 106; Conservative 6; Mismatches 17; Indels 2; Gaps 1;  
QY 1 QVQLQESGGGLVQAGGSLRLSCAASGTLITITMDVYRQTPGKQRELVAIGRTAGSTNYA 60  
DB 1 QVQLQESGGGLVQAGGSLRLSCAASGTLITITMDVYRQTPGKQRELVAIGRTAGSTNYA 60  
QY 61 DSAKGRFTISKDNAAKTVYLQWNSLKPEPTAVYYCNALITWDSS--VNDVWGSGTQVTV 118  
DB 61 DSAKGRFTISKDNAAKTVYLQWNSLKPEPTAVYYCNALITWDSS--VNDVWGSGTQVTV 120  
QY 119 SEPKTPKPOP 129  
DB 121 SEPKTPKPOP 131  
RESULT 4  
AAE10556  
ID AAE10556 standard; peptide; 130 AA.  
XX  
XX AAE10556;  
XX  
XX 10-DEC-2001 (first entry)  
XX  
XX HPL inhibiting VH fragment, HPL #18 from llama species.  
XX  
XX llama antibody; camelid; anorectic; heavy chain variable domain; VH;  
KW human dietary enzyme inhibitor; medicament; human pancreatic lipase; HPL;  
KW food; human gastric lipase; HGL; cosmetic control; body weight.  
XX  
XX Lama SP.  
XX  
XX Key Location/Qualifiers  
XX Region 31..35  
XX /label= CDR1  
XX /note= "Complementarity determining region 1"  
XX Region 50..64  
XX /label= CDR2  
XX /note= "Complementarity determining region 2"  
XX Region 98..110  
XX /label= CDR3  
XX /note= "Complementarity determining region 3"  
XX  
XX EPI134231-A1.  
XX  
XX 19-SEP-2001.  
XX  
XX 20-FEB-2001; 2001EP-00200703.  
XX

PR 14-MAR-2000; 2000EP-00200930.  
XX  
XX (UNIL ) UNILEVER NV.  
PA (UNIL ) UNILEVER P.C.  
XX  
XX Bezemer S, Van De Burg M, De Haard JWW, Tareilus E;  
XX WPI; 2001-572718/65.  
XX  
XX New antibody or its fragments for inhibiting human dietary enzymes,  
PT useful for cosmetic control of body weight of human beings, comprises  
PT heavy chain variable domain derived from immunoglobulin naturally devoid  
PT of light chains.  
XX  
XX Example 2; Page 10; 37pp; English.  
XX  
XX The patent discloses antibodies or their fragments comprising a heavy  
CC chain variable domain (VH) derived from an immunoglobulin naturally  
CC devoid of light chains specific for inhibiting human dietary enzymes. The  
CC antibodies of the invention are useful for the preparation of medicaments  
CC or food for inhibiting the activity of one or more human dietary enzymes  
CC especially human pancreatic lipase (HPL) or human gastric lipase (HGL)  
CC which are useful for the cosmetic control of body weight of human beings.  
CC The present peptide sequence is HPL inhibiting VH fragment, HPL #18 from  
CC llama (camelid) species  
XX  
XX Sequence 130 AA;  
SQ  
Query Match 76.4%; Score 518.5; DB 4; Length 130;  
Best Local Similarity 78.5%; Pred. No. 5.9e-38;  
Matches 102; Conservative 6; Mismatches 21; Indels 1; Gaps 1;  
QY 1 QVQLQESGGGLVQAGGSLRLSCAASGTLITITMDVYRQTPGKQRELVAIGRTAGSTNYA 60  
DB 1 QVQLQESGGGLVQAGGSLRLSCAASGTLITITMDVYRQTPGKQRELVAIGRTAGSTNYA 60  
QY 61 DSAKGRFTISKDNAAKTVYLQWNSLKPEPTAVYYCNALITR-WDKSVNYWGGTQVTVS 119  
DB 61 DSAKGRFTISKDNAAKTVYLQWNSLKPEPTAVYYCNALITR-WDKSVNYWGGTQVTVS 120  
QY 120 SEPKTPKPOP 129  
DB 121 SEPKTPKPOP 130  
RESULT 5  
AAE05283  
ID AAE05283 standard; protein; 152 AA.  
XX  
XX AAE05283;  
XX  
XX 18-SEP-2001 (first entry)  
XX  
XX Anti-potato SBII (Clone46) VH region attached with myc and his tag.  
XX  
XX potato; heavy chain immunoglobulin; pathogen resistance;  
KW metabolism modulator; passive immunisation; heavy chain variable domain;  
KW VH; anti-potato SBII; starch branching enzyme; SBE A.  
XX  
XX Solanum tuberosum.  
OS Undifferentiated.  
OS Chimeric.  
XX  
XX EPI118669-A2.  
XX  
XX 25-JUL-2001.  
XX  
XX 08-DEC-2000; 2000EP-00310997.  
XX  
XX 17-DEC-1999; 99EP-00310188.  
XX  
XX (UNIL ) UNILEVER P.C.  
XX (UNIL ) UNILEVER NV.  
XX

XX Frenken LGJ, Van Der Logt CPE, Jobling SA, Teh Y;  
 PI WPI; 2001-427157/46.  
 XX DR N-PSDB; AADI0054.  
 XX

PT Modifying a plant to produce an antibody useful for increasing pathogen  
 PT resistance or to modulate metabolism comprises introducing a DNA sequence  
 PT encoding a heavy chain immunoglobulin linked to a peptide that targets a  
 PT cellular compartment.

XX Example 1; Fig 12; 81pp; English.

XX The present invention relates to a method for modifying a plant to  
 CC produce an antibody or an active fragment or derivative, or a protein  
 CC functional equivalent, in a cellular compartment comprising introducing a  
 CC DNA sequence encoding a heavy chain immunoglobulin, where the DNA is  
 CC linked to promoters and provided with an additional sequence encoding a  
 CC peptide capable of targeting heavy chain immunoglobulin to a cellular  
 CC compartment. The method is used for producing a heavy chain  
 CC immunoglobulin or an active fragment or derivative, or a protein that is  
 CC functionally equivalent for increasing the pathogen resistance in a plant  
 CC or to modulate metabolism in a plant. Under some circumstances it may be  
 CC desirable to retain the antibody product with the plant rather than  
 CC extracting and isolating the product. In particular, edible selected  
 CC antigens may be used in a method of passively immunising an animal.  
 CC Preferably human, against the antigen, e.g., pathogenic organisms. The  
 CC present sequence is an anti-potato SBEII antibody (denoted Clone46) heavy  
 CC chain variable domain (VH) attached to peptide linkers, myc and his6 tag.  
 CC The potato SBEII is a starch branching enzyme also designated SBE A  
 XX

XX Sequence 152 AA;

Query Match 74.7%; Score 507.5; DB 4; Length 152;  
 Best Local Similarity 77.3%; Pred. No. 6.5e-37;

Matches 102; Conservative 5; Mismatches 16; Indels 9; Gaps 2;

QY 1 QVQLQESGGGLVQAGSRLRSCAASGTLISITIMDYRQTPGKQRELVGRITAGSSTNYA 60

DB 1 QVQLQESGGGLVQAGSRLRSCAASGTLISITIMDYRQTPGKQRELVGRITAGSSTNYA 60

QY 61 DSAKGRFTISKDNKNTVYIQNNSLKPEDTAVYYC---NALLTRMKSNVDYWGQGTQVTT 117

DB 61 DSVKGRFTISRDKAKNTVYIQNNSLKPEDTAVYYCAAGNLVYK-----PWGGQTLVLT 114

QY 118 VSSEPTPKPQP 129

DB 115 VSSEPTPKPQP 126

RESULT 6  
 AAE10557 standard; peptide; 130 AA.

XX AAE10557;

XX 10-DEC-2001 (first entry)

XX HPL inhibiting VHH fragment, HPL #19 from llama species.

XX llama antibody; camelid; anorectic; heavy chain variable domain; VHH;  
 KW human dietary enzyme inhibitor; medicament; human pancreatic lipase; HPL;  
 KW food; human gastric lipase; HGL; cosmetic control; body weight.

XX Lama sp.

XX Location/Qualifiers

XX Key 31..35

XX Region /label=CDR1

XX /note="Complementarity determining region 1"

XX /label=CDR2

XX /note="Complementarity determining region 2"

FT Region 98..110  
 FT /label=CDR3  
 FT /note="Complementarity determining region 3"

XX EP1134231-A1.

XX 19-SEP-2001.

XX 20-FEB-2001; 2001EP-00200703.

XX 14-MAR-2000; 2000EP-00200930.

XX (UNIL ) UNILEVER NV.

XX (UNIL ) UNILEVER PLC.

XX Bezemmer S, Van De Burg M, De Haard JWM, Tarellus E;

XX WPI; 2001-572718/65.

XX New antibody or its fragments for inhibiting human dietary enzymes,  
 PT useful for cosmetic control of body weight of human beings, comprises  
 PT heavy chain variable domain derived from immunoglobulin naturally devoid  
 PT of light chains.

XX Example 2; Page 10; 37pp; English.

XX The patent discloses antibodies or their fragments comprising a heavy  
 CC chain variable domain (VHH) derived from an immunoglobulin naturally  
 CC devoid of light chains specific for inhibiting human dietary enzymes. The  
 CC antibodies of the invention are useful for the preparation of medicaments  
 CC or food for inhibiting the activity of one or more human dietary enzymes  
 CC especially human pancreatic lipase (HPL) or human gastric lipase (HGL),  
 CC which are useful for the cosmetic control of body weight of human beings.  
 CC The present peptide sequence is HPL inhibiting VHH fragment, HPL #19 from  
 CC llama (camelid) species

XX Sequence 130 AA;

Query Match 74.4%; Score 505.5; DB 4; Length 130;  
 Best Local Similarity 74.6%; Pred. No. 8.2e-37;

Matches 97; Conservative 8; Mismatches 24; Indels 1; Gaps 1;

QY 1 QVQLQESGGGLVQAGSRLRSCAASGTLISITIMDYRQTPGKQRELVGRITAGSSTNYA 60

DB 1 QVQLQESGGGLVQAGSRLRSCAASGTLISITIMDYRQTPGKQRELVGRITAGSSTNYA 60

QY 61 DSAKGRFTISKDNKNTVYIQNNSLKPEDTAVYYCNALL-TRWDKSNVDYWGQGTQVTT 119

DB 61 DSVKGRFTISRDKAKNTVYIQNNSLKPEDTAVYYCNAQVFRSSDYTYWGQGTQVTT 120

QY 120 SEPTPKPQP 129

DB 121 SEPTPKPQP 130

RESULT 7

XX AAE10552 standard; peptide; 130 AA.

XX AAE10552;

XX 10-DEC-2001 (first entry)

XX HPL inhibiting VHH fragment, HPL #12 from llama species.

XX llama antibody; camelid; anorectic; heavy chain variable domain; VHH;  
 KW human dietary enzyme inhibitor; medicament; human pancreatic lipase; HPL;  
 KW food; human gastric lipase; HGL; cosmetic control; body weight.

XX Lama sp.

XX Location/Qualifiers

XX Key 31..35

XX Region



DE Anti-potato SBEII (Clone68) VH region attached with myc and his6 tag.  
 XX  
 XX Potato; heavy chain immunoglobulin; pathogen resistance;  
 KW metabolism modulator; passive immunisation; heavy chain variable domain;  
 KM VH; anti-potato SBEII; starch branching enzyme; SB2 A.  
 XX  
 OS Solanum tuberosum.  
 OS Unidentified.  
 OS Chimeric.  
 XX  
 XX BP118669-A2.  
 XX  
 XX 25-JUL-2001.  
 XX  
 XX 08-DEC-2000; 2000EP-00310997.  
 XX  
 XX 17-DEC-1999; 99EP-00310188.  
 XX  
 XX (UNIL ) UNILEVER PLC.  
 XX (UNIL ) UNILEVER NV.  
 XX  
 XX Freuden LGU, Van Der Logt CPE, Jobling SA, Teh Y;  
 DR WPI; 2001-427157/46.  
 DR N-PSDB; AADI0059.  
 XX  
 PT Modifying a plant to produce an antibody useful for increasing pathogen  
 PT resistance or to modulate metabolism comprises introducing a DNA sequence  
 PT encoding a heavy chain immunoglobulin linked to a peptide that targets a  
 PT cellular compartment.  
 XX  
 XX Example 13; Fig 28; 81pp; English.  
 XX  
 CC The present invention relates to a method for modifying a plant to  
 CC produce an antibody or an active fragment or derivative, or a protein  
 CC functional equivalent, in a cellular compartment comprising introducing a  
 CC DNA sequence encoding a heavy chain immunoglobulin; where the DNA is  
 CC linked to promoters and provided with an additional sequence encoding a  
 CC peptide capable of targeting heavy chain immunoglobulin to a cellular  
 CC compartment. The method is used for producing a heavy chain  
 CC immunoglobulin or an active fragment or derivative, or a protein that is  
 CC functionally equivalent for increasing the pathogen resistance in a plant  
 CC or to modulate metabolism in a plant. Under some circumstances it may be  
 CC desirable to retain the antibody product with the plant rather than  
 CC extracting and isolating the product. In particular, edible selected  
 CC antigens may be used in a method of passively immunising an animal. The  
 CC preferably human, against the antigen, e.g., pathogenic organisms. The  
 CC present sequence is an anti-potato SBEII antibody (denoted Clone68) heavy  
 CC chain variable domain (VH) attached to peptide linkers, myc and his6 tag.  
 CC The potato SBEII is a starch branching enzyme also designated SBE A  
 XX  
 XX Sequence 149 AA;  
 SQ  
 Query Match 73.0%; Score 496; DB 4; Length 149;  
 Best Local Similarity 75.2%; Pred. No. 6.5e-36;  
 Matches 97; Conservative 9; Mismatches 15; Indels 8; Gaps 1;

QY 1 QVOLOESGGGLVQAGSLRLSCAAGTILSIYDMWYRQTPGKORLVGRITAGGSTNYA 60  
 DB 3 QVOLOESGGGLVQAGSLRLSCAAGTILSIYDMWYRQTPGKORLVGRITAGGSTNYA 62  
 QY 61 DSAKGRFTISKDNARKNTVYLQNNSLKPEDTAVVVYCAALITPMKSNVDYWGQGTQVTVSS 120  
 DB 63 DSVKGRFTISKDNARKNTVYLQNNSLKPEDTAVVVYCAALITPMKSNVDYWGQGTQVTVSS 114  
 QY 121 EPKTPKPOP 129  
 DB 115 EPKTPKPOP 123

RESULT 10  
 ABG30620 standard; protein; 131 AA.  
 ID ABG30620 standard; protein; 131 AA.

XX  
 AC ABG30620;  
 XX  
 XX 21-OCT-2002 (first entry)  
 DT  
 XX  
 XX Immunoglobulin G specific heavy chain variable domain antibody #3.  
 DE  
 XX Heavy chain variable domain; antibody; protein array; aging; VH;  
 KM immunoglobulin; sticky phage-antibody; micro-panning; immune library;  
 KM proteomic; mouse.  
 XX  
 OS Mus sp.  
 OS  
 XX W0200248193-A2.  
 XX  
 XX 20-JUN-2002.  
 XX  
 XX 03-DEC-2001; 2001NO-EP014471.  
 XX  
 XX 13-DEC-2000; 2000EP-00311142.  
 XX  
 XX (UNIL ) UNILEVER NV.  
 XX (UNIL ) UNILEVER PLC.  
 XX (UNIL ) HINDUSTAN LEVER LTD.  
 XX  
 XX De Haard JWM, Hermans P, Landa I, Verrips CT;  
 DR WPI; 2002-583487/62.  
 XX  
 PT Novel protein array useful for detecting the presence of individual  
 PT proteins in sample, comprises heavy-chain variable domain antibodies or  
 PT antibody fragments obtainable from Camelidae.  
 XX  
 XX Disclosure; Fig 3; 80pp; English.  
 XX  
 CC The invention describes a protein array (I) comprising a number of heavy-  
 CC chain variable domain antibodies or antibody fragments, obtainable from  
 CC Camelidae. The method is useful for removing abundant proteins from an  
 CC extract or sample which do not provide useful information on the  
 CC condition of a cell or tissue in the extract or sample to be  
 CC investigated. (I) is useful for detecting the presence of individual  
 CC proteins in a sample, comparing the distribution of proteins in different  
 CC cell types, and identification of proteins that may be of importance in  
 CC determining the altered properties of cells in disease, aging or other  
 CC conditions. Using a heavy-chain variable domain derived from an  
 CC immunoglobulin that is naturally devoid of light chains (VHH) in (I)  
 CC provides a number of advantages, such as an improvement of  
 CC sensitivity/resolution in the order of 10-100 times, and detection of  
 CC post-translationally modified proteins. The invention also describes a  
 CC method (II) that enables the simultaneous processing of large numbers of  
 CC target antigens in a controlled way. The incorporated phage-ELISA  
 CC generates on-line information about the success or failure of a certain  
 CC panning condition. This feature combined with the microtiter plate format  
 CC allows the complete automation of the technology, based on computer-made  
 CC decisions on the values of the phage-ELISA for continuation of a limited  
 CC number of selections. In (II), due to the fact that many different  
 CC conditions can be tested, varying amounts of input-phages can be used  
 CC simultaneously in order to decrease the enrichment of sticky phage-  
 CC antibodies. Micro-panning is an effective tool for selecting both native,  
 CC synthetic and immune libraries on large panels of different target  
 CC molecules, enabling the generation of large panels of different target  
 CC time frames needed for the generation of arrays (proteomes). This  
 CC sequence represents an immunoglobulin (IgG) heavy chain variable domain  
 CC antibody (VHH)  
 XX  
 XX Sequence 131 AA;  
 SQ  
 Query Match 72.8%; Score 494; DB 5; Length 131;  
 Best Local Similarity 74.8%; Pred. No. 8.5e-36;  
 Matches 98; Conservative 6; Mismatches 25; Indels 2; Gaps 1;

QY 1 QVOLOESGGGLVQAGSLRLSCAAGTILSIYDMWYRQTPGKORLVGRITAGGSTNYA 60  
 ID QVOLOESGGGLVQAGSLRLSCAAGTILSIYDMWYRQTPGKORLVGRITAGGSTNYA 60

Db 1 QVQLQESGGGLVQPGGSLRLSCAASKSIFGCAVAMHROAPKQRELVATITDSTGNTYA 60  
 QY 61 DSAKGRFTISKDNAAKNTVYLQWNSLKPEDTAVYYCNALITR--MDKSYNDYWGQGTQVTV 118  
 Db 61 DSVKGRFTISRARNKNTVYLQWNSLKPEDTGYVYCNASTVATITGRFTITDLMGQGTQVTV 120  
 QY 119 SESEPTPKPQP 129  
 Db 121 SESEPTPKPQP 131

## RESULT 11

AAE10561  
 ID AAE10561 standard; peptide; 129 AA.

AC AAE10561;  
 DT 10-DEC-2001 (first entry)  
 XX HGL inhibiting VHH fragment, HGL #4 from llama species.  
 DE HGL inhibiting VHH fragment, HGL #4 from llama species.  
 XX Llama antibody; camelid; anorectic; heavy chain variable domain; VHH;  
 KW human dietary enzyme inhibitor; medicament; human pancreatic lipase; HPL;  
 KM food; human gastric lipase; HGL; cosmetic control; body weight.  
 XX  
 OS Lama sp.  
 XX  
 FH Key Location/Qualifiers  
 FT Region 31..35  
 FT /label=CDR1  
 FT /note="Complementarity determining region 1"  
 FT 50..64  
 FT /label=CDR2  
 FT /note="Complementarity determining region 2"  
 FT 98..109  
 FT /label=CDR3  
 FT /note="Complementarity determining region 3"

EP1134231-A1.

19-SEP-2001.

20-FEB-2001; 2001EP-00200703.

14-MAR-2000; 2000EP-00200930.

PA (UNIL ) UNILEVER NV.

PA (UNIL ) UNILEVER PLC.

PI Bezemer S, Van De Burg M, De Haard JWW, Tareilus E;

WPI; 2001-572718/65.

PT New antibody or its fragments for inhibiting human dietary enzymes,  
 PT useful for cosmetic control of body weight of human beings, comprises  
 PT heavy chain variable domain derived from immunoglobulin naturally devoid  
 PT of light chains.

Example 4; Page 13; 37pp; English.

XX The patent discloses antibodies or their fragments comprising a heavy  
 CC chain variable domain (VHH) derived from an immunoglobulin naturally  
 CC devoid of light chains specific for inhibiting human dietary enzymes. The  
 CC antibodies of the invention are useful for the preparation of medicaments  
 CC or food for inhibiting the activity of one or more human dietary enzymes  
 CC especially human pancreatic lipase (HPL) or human gastric lipase (HGL)  
 CC which are useful for the cosmetic control of body weight of human beings.  
 CC The present peptide sequence is HGL inhibiting VHH fragment, HGL #4 from  
 CC llama (camelid) species  
 XX  
 XX Sequence 129 AA;

Query Match 72.6%; Score 493; DB 4; Length 129;

Best Local Similarity 75.2%; Pred. No. 1e-35;  
 Matches 97; Conservative 7; Mismatches 25; Indels 0; Gaps 0;

QY 1 QVQLQESGGGLVQAGSLRLSCAASGTTLSIIMDWYRCPKQRELVATITGSGSTNYA 60  
 Db 1 QVQLQESGGGLVQAGSLRLSCAASGTFEFNMGWYRCPKQRELVAAIGDGLTYY 60  
 QY 61 DSAKGRFTISKDNAAKNTVYLQWNSLKPEDTAVYYCNALITRMDKSYNDYWGQGTQVTVSS 120  
 Db 61 DSVKGRFTISRARNKNTVYLQWNSLKPEDTAVYYCKRGGLTQYSEHDYWGQGTQVTVSS 120  
 QY 121 EPKTPKQP 129  
 Db 121 EPKTPKQP 129

## RESULT 12

AAE10554  
 ID AAE10554 standard; peptide; 130 AA.

AC AAE10554;  
 DT 10-DEC-2001 (first entry)  
 XX HPL inhibiting VHH fragment, HPL #14 from llama species.  
 DE HPL inhibiting VHH fragment, HPL #14 from llama species.  
 XX Llama antibody; camelid; anorectic; heavy chain variable domain; VHH;  
 KW human dietary enzyme inhibitor; medicament; human pancreatic lipase; HPL;  
 KM food; human gastric lipase; HGL; cosmetic control; body weight.  
 XX  
 OS Lama sp.

FH Key Location/Qualifiers  
 FT Region 31..35  
 FT /label=CDR1  
 FT /note="Complementarity determining region 1"  
 FT 50..64  
 FT /label=CDR2  
 FT /note="Complementarity determining region 2"  
 FT 98..110  
 FT /label=CDR3  
 FT /note="Complementarity determining region 3"

EP1134231-A1.

19-SEP-2001.

20-FEB-2001; 2001EP-00200703.

14-MAR-2000; 2000EP-00200930.

PA (UNIL ) UNILEVER NV.

PA (UNIL ) UNILEVER PLC.

PI Bezemer S, Van De Burg M, De Haard JWW, Tareilus E;

WPI; 2001-572718/65.

PT New antibody or its fragments for inhibiting human dietary enzymes,  
 PT useful for cosmetic control of body weight of human beings, comprises  
 PT heavy chain variable domain derived from immunoglobulin naturally devoid  
 PT of light chains.

Example 2; Page 9; 37pp; English.

XX The patent discloses antibodies or their fragments comprising a heavy  
 CC chain variable domain (VHH) derived from an immunoglobulin naturally  
 CC devoid of light chains specific for inhibiting human dietary enzymes. The  
 CC antibodies of the invention are useful for the preparation of medicaments  
 CC or food for inhibiting the activity of one or more human dietary enzymes  
 CC especially human pancreatic lipase (HPL) or human gastric lipase (HGL)  
 CC which are useful for the cosmetic control of body weight of human beings.  
 CC The present peptide sequence is HPL inhibiting VHH fragment, HPL #14 from

CC llama (camelid) species  
XX  
XX Sequence 130 AA;

Query Match 72.5%; Score 492.5; DB 4; Length 130;  
Best Local Similarity 76.2%; Pred. No. 1.1e-35;  
Matches 99; Conservative 9; Mismatches 21; Indels 1; Gaps 1;

QY 1 QVQLQESGGGLVQAGSLRLSCAASGTLISITIMDWRQTPGKRELVGRITAGSTNYA 60  
DB 1 QVQLQESGGGLVQAGSLRLSCAASGTLISITIMDWRQTPGKRELVGRITAGSTNYA 60  
QY 61 DSAKGRFTISRDNAKNTVYLQNNSLKPEPTAVYYCAALTRPDKS-VNDYWGQGTQVTVS 119  
DB 61 DSAKGRFTISRDNIINTVYLQNNSLKPEPTAVYHCNADYPRYRTRYLELWGQGTQVTVS 120  
QY 120 SEPKTPKPOP 129  
DB 121 SEPKTPKPOP 130

## RESULT 13

AAE10563  
ID AAE10563 standard; peptide; 124 AA.

AC AAE10563;  
XX  
XX 10-DEC-2001 (first entry)  
DE HGL inhibiting VHH fragment, HGL #9 from llama species.  
XX

XX llama antibody; camelid; anorectic; heavy chain variable domain; VHH;  
XX human dietary enzyme inhibitor; medicament; human pancreatic lipase; HPL;  
XX food; human gastric lipase; HGL; cosmetic control; body weight.

OS Lama sp.  
XX  
XX Key Location/Qualifiers  
FH 31..35  
FT Region /label= CDR1  
FT /note= "Complementarity determining region 1"  
FT Region 50..64  
FT /label= CDR2  
FT /note= "Complementarity determining region 2"  
FT Region 98..104  
FT /label= CDR3  
FT /note= "Complementarity determining region 3"

PN EPI134231-A1.  
XX  
XX 19-SEP-2001.  
PD  
XX 20-FEB-2001; 2001EP-00200703.  
PF  
XX 14-MAR-2000; 2000EP-00200930.  
PR

XX (UNITL ) UNILEVER NV.  
PA  
XX (UNITL ) UNILEVER PLC.  
PI

XX Bezemer S, Van De Burg M, De Haard JWM, Tareilus E;  
PI  
XX WPI; 2001-572718/65.  
DR

XX New antibody or its fragments for inhibiting human dietary enzymes,  
PT useful for cosmetic control of body weight of human beings, comprises  
PT heavy chain variable domain derived from immunoglobulin naturally devoid  
PT of light chains.

XX Example 4; Page 13; 37pp; English.

XX The patent discloses antibodies or their fragments comprising a heavy  
CC chain variable domain (VHH) derived from an immunoglobulin naturally  
CC devoid of light chains specific for inhibiting human dietary enzymes. The

CC antibodies of the invention are useful for the preparation of medicaments  
CC or food for inhibiting the activity of one or more human dietary enzymes  
CC especially human pancreatic lipase (HPL) or human gastric lipase (HGL)  
CC which are useful for the cosmetic control of body weight of human beings.  
CC The present peptide sequence is HGL inhibiting VHH fragment, HGL #9 from  
CC llama (camelid) species  
XX  
XX Sequence 124 AA;

Query Match 72.1%; Score 489.5; DB 4; Length 124;  
Best Local Similarity 74.4%; Pred. No. 2e-35; Indels 5; Gaps 1;  
Matches 96; Conservative 6; Mismatches 22;

QY 1 QVQLQESGGGLVQAGSLRLSCAASGTLISITIMDWRQTPGKRELVGRITAGSTNYA 60  
DB 1 QVQLQESGGGLVQAGSLRLSCAASGTLISITIMDWRQTPGKRELVGRITAGSTNYA 60  
QY 61 DSAKGRFTISRDNAKNTVYLQNNSLKPEPTAVYYCAALTRPDKS-VNDYWGQGTQVTVS 120  
DB 61 DSAKGRFTISRDNIKTMYLQNNSLKPEPTGVYCAG-----TGAEGHYWGQGTQVTVS 115  
QY 121 EPKTPKPOP 129  
DB 116 EPKTPKPOP 124

## RESULT 14

AAE10564  
ID AAE10564 standard; peptide; 129 AA.

AC AAE10564;  
XX  
XX 10-DEC-2001 (first entry)  
DE HGL inhibiting VHH fragment, HGL #10 from llama species.  
XX

XX llama antibody; camelid; anorectic; heavy chain variable domain; VHH;  
XX human dietary enzyme inhibitor; medicament; human pancreatic lipase; HPL;  
XX food; human gastric lipase; HGL; cosmetic control; body weight.

OS Lama sp.  
XX  
XX Key Location/Qualifiers  
FH 31..35  
FT Region /label= CDR1  
FT /note= "Complementarity determining region 1"  
FT Region 50..64  
FT /label= CDR2  
FT /note= "Complementarity determining region 2"  
FT Region 98..109  
FT /label= CDR3  
FT /note= "Complementarity determining region 3"

PN EPI134231-A1.  
XX  
XX 19-SEP-2001.  
PD  
XX 20-FEB-2001; 2001EP-00200703.  
PF  
XX 14-MAR-2000; 2000EP-00200930.  
PR

XX (UNITL ) UNILEVER NV.  
PA  
XX (UNITL ) UNILEVER PLC.  
PI

XX Bezemer S, Van De Burg M, De Haard JWM, Tareilus E;  
PI  
XX WPI; 2001-572718/65.  
DR

XX New antibody or its fragments for inhibiting human dietary enzymes,  
PT useful for cosmetic control of body weight of human beings, comprises  
PT heavy chain variable domain derived from immunoglobulin naturally devoid  
PT of light chains.

PS Example 4; Page 13; 37pp; English.

The patent discloses antibodies or their fragments comprising a heavy chain variable domain (V<sub>H</sub>) derived from an immunoglobulin naturally devoid of light chains specific for inhibiting human dietary enzymes. The antibodies of the invention are useful for the preparation of medicaments or food for inhibiting the activity of one or more human dietary enzymes especially human pancreatic lipase (HPL) or human gastric lipase (HGL) which are useful for the cosmetic control of body weight of human beings. The present peptide sequence is HGL inhibiting V<sub>H</sub> fragment, HGL #10 from llama (camelid) species

Sequence 129 AA;

Query Match	72.0%;	Score 489;	DB 4;	Length 129;
Best Local Similarity	75.2%;	Pred. No. 2.3e-35;		

QY 1 QVQLQESGGGLVQAGGSLRSCAASGTILSIYMDWYRQTPGKQRELVGRI TAGGSTNYA 60

LD 1 QVQLQESGGDLVQAAGSLRLACASGSTFSFNAMGMYRQVPGKQRELVAIGNDGITYV 60

61 DSAGKRF11SKUNAKNI-VLLOJNSLKPEJIAVYYCNALITRMDKS VNDYWGQGTQTVSS 120

DD 01 NSVNGRT I SKRENANNI VILQUNSLNRPEDIAV I CNGRGGLQISEHDIWQDGIQVIVSS 120

[illegible]

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
84

## RESULT 15

AAE10567 standard; peptide; 130 AA.

DT 10-DEC-2001 (first entry)  
YY

HGL inhibiting VHH fragment, HGL #16 from llama species.

KW llama antibody; camelid; anorectic; heavy chain variable domain; VHH;  
KW human dietary enzyme inhibitor; medicament; human pancreatic lipase; HPL  
KW food; human gastric lipase; HGL; cosmetic control; body weight.

OS Lama sp  
vxy

EH	Key	Location/Qualifiers
ET		31 35

FT	Region	31.	.35
----	--------	-----	-----

/note= "Complementarity determining region 1"

```

E1 /label= CDR2
E1 /noconcat=0;

```

FT	Region	98.110	label = CDR3
----	--------	--------	--------------

```

F.1 /note= "Complementarity determining region 3"
XX

```

XX	1
XX	2
XX	3
XX	4
XX	5
XX	6
XX	7
XX	8
XX	9
XX	10
XX	11
XX	12
XX	13
XX	14
XX	15
XX	16
XX	17
XX	18
XX	19
XX	20
XX	21
XX	22
XX	23
XX	24
XX	25
XX	26
XX	27
XX	28
XX	29
XX	30
XX	31
XX	32
XX	33
XX	34
XX	35
XX	36
XX	37
XX	38
XX	39
XX	40
XX	41
XX	42
XX	43
XX	44
XX	45
XX	46
XX	47
XX	48
XX	49
XX	50
XX	51
XX	52
XX	53
XX	54
XX	55
XX	56
XX	57
XX	58
XX	59
XX	60
XX	61
XX	62
XX	63
XX	64
XX	65
XX	66
XX	67
XX	68
XX	69
XX	70
XX	71
XX	72
XX	73
XX	74
XX	75
XX	76
XX	77
XX	78
XX	79
XX	80
XX	81
XX	82
XX	83
XX	84
XX	85
XX	86
XX	87
XX	88
XX	89
XX	90
XX	91
XX	92
XX	93
XX	94
XX	95
XX	96
XX	97
XX	98
XX	99
XX	100

XX

XX

XX XX  
XX XX

PA (UNIL ) UNILEVER PLC.  
XX

Bezemer S, Van De Burg M, De Haard JWM, Tarelilus E;  
XX

NET / NOV - 3 / 2 / 10 / 03.  
XX  
XX

PT New antibody or its fragments for inhibiting human dietary enzymes,  
PT useful for cosmetic control of body weight of human beings, comprises  
PT heavy chain variable domain derived from immunoglobulin naturally devoid  
PT of light chains.

Example 4; page 14; 37pp; English.

CC The patent discloses antibodies or their fragments comprising a heavy  
CC chain variable domain (V<sub>H</sub>H) derived from an immunoglobulin naturally  
CC devoid of light chains specific for inhibiting human dietary enzymes. The  
CC antibodies of the invention are useful for the preparation of medicaments  
CC or food for inhibiting the activity of one or more human dietary enzymes  
CC especially human pancreatic lipase (HPL) or human gastric lipase (HGL)  
CC which are useful for the cosmetic control of body weight of human beings.  
CC The present peptide sequence is HGL inhibiting V<sub>H</sub>H fragment, HGL #16 from  
CC llama (camelid) species

Sequence 130 AA

Query Match	71.2%;	Score 483.5;	DB 4;	Length 130;
Best Local Similarity	73.3%;	Pred. No. 7.1e-35;		
Matches 99;	Conservative 6;	Mismatches 19;	Indels 11;	Gaps 2

QY 1 QVQLQESGGGLVQAGGSLRLSCAASGTTLSIIYMDWYRQTPGKQRELVGRTAGGSTNYA 60

DB 1 QVQLQESGGGLVQAGGSLKLSCAASGSDFRYNAMAWYRQAPGKQRLVAITITIKRTNYA 60

01 DOMINANT 110NENPNU I VI E QWBNED IAWI I CUNALI INNDNSVN --- D1M9Y01 11

[illegible][illegible]

```
Search completed: October 5, 2004, 08:00:10
Job time : 98.1435 secs
```

This page blank (uspo)



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: October 5, 2004, 07:36:16 ; Search time 21.0648 Seconds  
(without alignments)  
593.639 Million cell updates/sec

Title: US-09-805-290a-21

Perfect score: 683  
Sequence: 1 QVQLQESGGGLVQAGGSLRLT.....WGQGLTVTSSEPKPKPPCP 130

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :  
1: Dirl:\*  
2: Pir2:\*  
3: Pir3:\*  
4: Pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	378	55.3	119	2	Ig heavy chain - h
2	376	55.1	140	2	Ig heavy chain V r
3	375	55.9	119	2	Ig heavy chain - h
4	374.5	54.8	120	2	Ig heavy chain V r
5	371	54.3	138	2	Ig heavy chain V r
6	368	53.9	121	2	Ig heavy chain V r
7	368	53.9	121	2	Ig heavy chain - h
8	368	53.9	123	2	Ig heavy chain - h
9	368	53.9	160	2	Ig heavy chain pre
10	365.5	53.7	151	2	Ig heavy chain pre
11	365.5	53.5	147	2	Ig heavy chain pre
12	365	53.4	119	2	Ig variable region
13	365	53.4	119	2	Ig heavy chain V r
14	363	53.1	123	2	Ig heavy chain V r
15	363	53.1	125	2	Ig heavy chain V r
16	362.5	53.1	116	2	Ig heavy chain V r
17	362.5	53.1	124	2	Ig heavy chain - h
18	362	53.0	120	2	Ig heavy chain V r
19	362	53.0	122	2	Ig heavy chain V r
20	362	53.0	127	2	Ig heavy chain V r
21	361.5	52.9	114	2	Ig heavy chain - h
22	361.5	52.9	140	2	Ig heavy chain V r
23	361	52.9	140	2	Ig heavy chain pre
24	360.5	52.8	120	2	Ig heavy chain V-D
25	359	52.6	117	2	Ig heavy chain V r
26	359	52.6	119	2	Ig heavy chain V r
27	359	52.6	127	2	Ig heavy chain V r
28	358.5	52.5	128	2	Ig heavy chain V r
29	358	52.4	135	2	Ig heavy chain V r

30	358	52.4	140	2	Ig heavy chain V r
31	357.5	52.3	128	2	Ig heavy chain V r
32	357	52.3	139	2	Ig heavy chain V r
33	356	52.1	115	1	Ig heavy chain V-L
34	356	52.1	117	2	Ig heavy chain - h
35	356	52.1	121	2	Ig heavy chain V r
36	355.5	52.0	119	2	Ig heavy chain V r
37	355	52.0	120	2	Ig heavy chain V r
38	355	52.0	134	2	Ig heavy chain V r
39	355	52.0	138	2	Ig heavy chain pre
40	354.5	51.9	128	2	Ig heavy chain V r
41	354	51.8	120	2	Ig heavy chain V r
42	354	51.8	122	2	Ig heavy chain V r
43	353.5	51.8	122	2	Ig heavy chain V r
44	353.5	51.8	122	2	Ig heavy chain - h
45	353	51.7	138	2	Ig heavy chain V r

## ALIGNMENTS

## RESULT 1

S31107  
Ig heavy chain - human  
C/Species: Homo sapiens (man)  
C/Date: 02-Dec-1993 #sequence\_revision 26-May-1995 #ext\_change 17-Mar-1999  
C/Accession: S31107  
R/Raaphorst, F.M.; Timmers, E.; Kenter, M.J.H.; van Tol, M.J.D.; Vossen, J.M.; Schuurman, Eur. J. Immunol. 22, 247-251, 1992  
A/Title: Restricted utilization of germ-line V(H)3 genes and short diverse third comple  
A/Reference number: S31104; MID:9211633; PMID:1730252  
A/Accession: S31107  
A/Status: preliminary; nucleic acid sequence not shown; translation not shown  
A/Molecule type: mRNA  
A/Residues: 1-119 <RA>  
A/Cross-references: EMBL:X62955  
A/Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1991  
C/Superfamily: immunoglobulin V region; immunoglobulin homology  
C/Keywords: heterotetramer; immunoglobulin  
F/15-98/Domain: immunoglobulin homology <IMM>

Query Match 55.3%; Score 378; DB 2; Length 119;  
Best Local Similarity 66.4%; Pred. No. 2e-26;  
Matches 81; Conservative 8; Mismatches 29; Indels 4; Gaps 2;

QY 1 QVQLQESGGGLVQAGGSLRLTSCAASGISIHMGWYRQTPGTERDVATIQ-DGSGTNY 59  
Db 1 EVQLVESGGGLVQPGGSLRLTSCAASGFTFSYAMSWVRQAPGKGLWVAISGGSGTNY 60  
QY 60 ADSVKGRFTISRNINLTNYVIONNSIKPEDTAVYHGNADVRPRTSRYLEMGQGLTVTV 119  
Db 61 ADSVKGRFTISRNINLTNYVIONNSIKPEDTAVYHGNADVRPRTSRYLEMGQGLTVTV 117  
QY 120 SS 121  
Db 118 SS 119

## RESULT 2

S31588  
Ig heavy chain V region - human (fragment)  
C/Species: Homo sapiens (man)  
C/Date: 22-Nov-1993 #sequence\_revision 10-Nov-1995 #ext\_change 23-Jul-1999  
C/Accession: S31588  
R/Christner, A.M.; Gauthier, L.; Boubli, L.; Fougereau, M.; Tonnelie, C.  
submitted to the EMBL Data Library, June 1992  
A/Description: Mechanisms that generate human immunoglobulin diversity operate from the  
A/Reference number: S31585  
A/Accession: S31588  
A/Status: preliminary  
A/Molecule type: mRNA  
A/Residues: 1-140 <CUI>  
A/Cross-references: EMBL:Z14200; NID:G30957; PIDN:CAA78569.1; PID:G30958



C:Keywords: heterotetramer; immunoglobulin  
F:15-98/Domain: immunoglobulin homology <IMM>

Query Match 53.9%; Score 368; DB 2; Length 121;  
Best Local Similarity 65.6%; Pred. No. 1.5e-25;  
Matches 80; Conservative 11; Mismatches 29; Indels 2; Gaps 2;

QY 1 QVQLQESGGGLVQAGGSLRLSCAASGSIQHTMGVYRQTPGERDVATIQDGS-TNY 59  
D 1 EVQLVESGGGLVQPGGSLRLSCTASGFNPSVHMWVRQAPGKGLVWYSHIRKGGITVNY 60

QY 60 ADSVKGRTISRDNINLTAVYLOMNSLKPEDTAVYHCNADVPYRTRSLYELMGQGLTVV 119  
D 61 ADSVKGRTISRDNKNTLYLOMNSLRBDTAVYVCVAIRGY-SYGLDLYMGQGLTVV 119

QY 120 SS 121  
D 120 SS 121

RESULT 7  
Ig heavy chain - human (fragment)  
C:Species: Homo sapiens (man)  
C:Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 21-Jan-2000  
C:Accession: 155673  
R:Knight, G.B.; Agnello, V.; Bonagura, V.; Barnes, J.L.; Panke, D.J.; Zhang, Q.X.  
J. Exp. Med. 178, 1903-1911, 1993  
A:Title: Human rheumatoid factor cross-idiotypes. IV. Studies on WA Xid-positive IGM with  
clint from the 17.109 and G6 Xids.  
A:Reference number: 155673; PMID:94065558; PMID:8245772  
A:Accession: 155673  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-121 <RES>  
A:Cross-references: GB:M87268; NID:G186197; PIDN:AAC37536.1; PID:G186198  
C:Gene: IGHM  
A:Gene: IGHM  
A:Cross-references: GDB:120086; OMIM:147020  
A:Map position: 14q32.33-14q32.33  
C:Superfamily: immunoglobulin V region; immunoglobulin homology  
F:15-98/Domain: immunoglobulin homology <IMM>

Query Match 53.9%; Score 368; DB 2; Length 121;  
Best Local Similarity 65.6%; Pred. No. 1.5e-25;  
Matches 80; Conservative 9; Mismatches 31; Indels 2; Gaps 2;

QY 1 QVQLQESGGGLVQAGGSLRLSCAASGSIQHTMGVYRQTPGERDVATIQDGS-TNY 59  
D 1 EVQLVESGGGLVQPGGSLRLSCTASGFNPSVHMWVRQAPGKGLVWYSHIRKGGITVNY 60

QY 60 ADSVKGRTISRDNINLTAVYLOMNSLKPEDTAVYHCNADVPYRTRSLYELMGQGLTVV 119  
D 61 ADSVKGRTISRDNKNTLYLOMNSLRBDTAVYVCVAIRGY-SYGLDLYMGQGLTVV 119

QY 120 SS 121  
D 120 SS 121

RESULT 8  
Ig heavy chain - human  
C:Species: Homo sapiens (man)  
C:Date: 02-Dec-1993 #sequence\_revision 26-May-1995 #text\_change 17-Mar-1999  
C:Accession: S31114  
R:Raaphorst, F.M.; Timmers, E.; Kenter, M.J.H.; van Tol, M.J.D.; Vossen, J.M.; Schuurmat  
Eur. J. Immunol. 22, 247-251, 1992  
A:Title: Restricted utilization of germ-line V(H)3 genes and short diverse third comple  
A:Reference number: S31104; PMID:92111633; PMID:1730252  
A:Accession: S31114  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: mRNA

A:Residues: 1-123 <RAA>  
A:Cross-references: EMBL:X62963  
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1991  
C:Superfamily: immunoglobulin V region; immunoglobulin homology  
C:Keywords: heterotetramer; immunoglobulin  
F:15-98/Domain: immunoglobulin homology <IMM>

Query Match 53.9%; Score 368; DB 2; Length 123;  
Best Local Similarity 62.8%; Pred. No. 1.6e-25;  
Matches 81; Conservative 9; Mismatches 25; Indels 14; Gaps 3;

QY 1 QVQLQESGGGLVQAGGSLRLSCAASGSIQHTMGVYRQTPGERDVATIQDGS-TNY 59  
D 1 EVQLVESGGGLVQPGGSLRLSCTASGFNPSVHMWVRQAPGKGLVWYSHIRKGGITVNY 60

QY 60 ADSVKGRTISRDNINLTAVYLOMNSLKPEDTAVYHCNADVPYRTRSLYELMGQGLTVV 112  
D 61 ADSVKGRTISRDNKNTLYLOMNSLRBDTAVYVCVAIRGY-SYGLDLYMGQGLTVV 114

QY 113 QGTLTVSS 121  
D 115 QGTLTVSS 123

RESULT 9  
Ig heavy chain precursor - human (fragment)  
C:Species: Homo sapiens (man)  
C:Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 16-Aug-1996  
C:Accession: S05271; S04602  
R:Kishimoto, T.  
submitted to the EMBL Data Library, March 1989  
A:Reference number: S05270  
A:Accession: S05271  
A:Molecule type: mRNA  
A:Residues: 1-160 <RISL>  
A:Cross-references: EMBL:X14584  
R:Kishimoto, T.; Okajima, H.; Okumoto, T.; Taniguchi, M.  
Nucleic Acids Res. 17, 4385, 1989  
A:Title: Nucleotide sequences of the cDNAs encoding the V-regions of H- and L-chains of  
A:Reference number: S04601; PMID:89256497; PMID:2500644  
A:Accession: S04602  
A:Molecule type: mRNA  
A:Residues: 1-144 <RIS2>  
A:Cross-references: EMBL:X14584  
C:Superfamily: immunoglobulin V region; immunoglobulin homology  
C:Keywords: heterotetramer; immunoglobulin  
F:1-19/Domain: signal sequence #status predicted <SIG>  
F:20-160/Product: Ig heavy chain (fragment) #status predicted <MAT>  
F:34-117/Domain: immunoglobulin homology <IMM>

Query Match 53.9%; Score 368; DB 2; Length 160;  
Best Local Similarity 64.8%; Pred. No. 2.1e-25;  
Matches 81; Conservative 11; Mismatches 29; Indels 4; Gaps 3;

QY 1 QVQLQESGGGLVQAGGSLRLSCAASGSIQHTMGVYRQTPGERDVATIQDGS-TNY 59  
D 20 EVQLVESGGGLVQPGGSLRLSCTASGFNPSVHMWVRQAPGKGLVWYSHIRKGGITVNY 79

QY 60 ADSVKGRTISRDNINLTAVYLOMNSLKPEDTAVYHCNADVPYRTRSLYELMGQGLTVV 116  
D 80 ADSVKGRTISRDNKNTLYLOMNSLRBDTAVYVCVAIRGY-SYGLDLYMGQGLTVV 139

QY 117 VTVSS 121  
D 140 VTVSS 144

RESULT 10  
Ig heavy chain precursor V region (clone HN.14) - human  
C:Species: Homo sapiens (man)  
C:Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 17-Mar-1999

C:Accession: A60943; A48165  
 J:Desai, R.; Spatz, L.; Matsuda, T.; Ilyas, A.A.; Berman, J.E.; Alt, F.W.; Kabat, E.A.;  
 J:Neuroimmunol 30, 245, 1990  
 A:Title: Molecular cloning of a human immunoglobulin heavy chain variable (V-H) region  
 C:Date: 16-Feb-1996 #sequence\_revision 16-Feb-1996 #text\_change 23-Jul-1999  
 A:Reference number: A60943; PMID:91036050; PMID:1699976  
 A:Accession: A60943  
 A:Status: not compared with conceptual translation  
 A:Molecule type: mRNA  
 A:Residues: 1-151 <DES>  
 J:Desai, R.; Spatz, L.; Matsuda, T.; Ilyas, A.A.; Berman, J.E.; Alt, F.W.; Kabat, E.A.;  
 J:Neuroimmunol 26, 35-41, 1990  
 A:Title: Molecular cloning of a human immunoglobulin heavy chain variable (V-H) region  
 A:Reference number: A48165; PMID:90094677; PMID:1688442  
 A:Accession: A48165  
 A:Molecule type: mRNA  
 A:Residues: 1-36, 'M', 38-62, 'AS', 67-151 <DE2>  
 A:Note: this sequence has been corrected in reference A60943  
 C:Superfamily: immunoglobulin V region; immunoglobulin homology  
 C:Keywords: heterotetramer; immunoglobulin  
 F:15-98/Domain: immunoglobulin homology <IMM>

Query Match 53.7%; Score 366.5; DB 2; Length 151;  
 Best Local Similarity 63.7%; Pred. No. 2.6e-25;  
 Matches 79; Conservative 10; Mismatches 32; Indels 3; Gaps 2;

OY 1 QVQLSSEGGGLVQAGSLRLSCAASGSIHTMGYRQTPGERDVATIQ-DGGSITNY 59  
 DB 20 QVQLVSGGGLVQPGSLRLSCAASGFTFSRYGMHWRAQPGGLEWVSISFDGITYY 79  
 OY 60 ADSVKGFTISRDNINLTNYVLOMNSLKPEPTAVYHCNADVRPYRTSRYEL--WGQGLTV 117  
 DB 80 ADSVKGFTISRDNINLTNYVLOMNSLKPEPTAVYHCNADVRPYRTSRYEL--WGQGLTV 139  
 OY 118 TVSS 121  
 DB 140 TVSS 143

RESULT 11  
 137780  
 IG variable region (VDJ) (clone T20-11) - human (fragment)  
 C:Species: Homo sapiens (man)  
 C:Date: 16-Feb-1996 #sequence\_revision 16-Feb-1996 #text\_change 23-Jul-1999  
 C:Accession: I37780; S25474  
 R:Demaison, C.; Chastagner, P.; Theze, J.; Zouali, M.  
 Proc. Natl. Acad. Sci. U.S.A. 91, 514-518, 1994  
 A:Title: Somatic diversification in the heavy chain variable region genes expressed by B  
 A:Reference number: A36876; PMID:9411917; PMID:8290556  
 A:Accession: I37780  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-147 <RES>  
 A:Cross-references: EMBL:X67943; NID:g33578; PIDD:CAA48130.1; PIDD:g33579  
 C:Superfamily: immunoglobulin V region; immunoglobulin homology  
 F:28-111/Domain: immunoglobulin homology <IMM>

Query Match 53.5%; Score 365.5; DB 2; Length 147;  
 Best Local Similarity 63.7%; Pred. No. 3.1e-25;  
 Matches 79; Conservative 10; Mismatches 32; Indels 3; Gaps 2;

OY 1 QVQLSSEGGGLVQAGSLRLSCAASGSIHTMGYRQTPGERDVATIQ-DGGSITNY 59  
 DB 14 EVQLVDSGGGLVQPGSLRLSCAASGFTFSYMSWVRQAPGKLEWVANIQDSGEKYY 73  
 OY 60 ADSVKGFTISRDNINLTNYVLOMNSLKPEPTAVYHCNADVRPYRTSRY--LELMGGGLTV 117  
 DB 74 ADSVKGFTISRDNINLTNYVLOMNSLKPEPTAVYHCNADVRPYRTSRY--LELMGGGLTV 133  
 OY 118 TVSS 121  
 DB 134 TVSS 137

RESULT 12  
 C36005  
 IG heavy chain V region (30P1) - human  
 C:Species: Homo sapiens (man)  
 C:Date: 21-Dec-1990 #sequence\_revision 21-Dec-1990 #text\_change 16-Aug-1996  
 A:Accession: C36005  
 R:Schroeder Jr., H.W.; Wang, J.Y.  
 Proc. Natl. Acad. Sci. U.S.A. 87, 6146-6150, 1990  
 A:Title: Preferential utilization of conserved immunoglobulin heavy chain variable gene  
 A:Reference number: A36005; PMID:90349571; PMID:2117273  
 A:Accession: C36005  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-119 <SCH>  
 A:Cross-references: GB:M18513  
 C:Superfamily: immunoglobulin V region; immunoglobulin homology  
 C:Keywords: heterotetramer; immunoglobulin  
 F:15-98/Domain: immunoglobulin homology <IMM>

Query Match 53.4%; Score 365; DB 2; Length 119;  
 Best Local Similarity 64.8%; Pred. No. 2.8e-25;  
 Matches 79; Conservative 8; Mismatches 31; Indels 4; Gaps 2;

OY 1 QVQLSSEGGGLVQAGSLRLSCAASGSIHTMGYRQTPGERDVATIQ-DGGSITNY 59  
 DB 1 EVQLSSEGGGLVQPGSLRLSCAASGFTFSYMSWVRQAPGKLEWVAISGSGSTYY 60  
 OY 60 ADSVKGFTISRDNINLTNYVLOMNSLKPEPTAVYHCNADVRPYRTSRYEL--WGQGLTV 119  
 DB 61 ADSVKGFTISRDNINLTNYVLOMNSLKPEPTAVYHCNADVRPYRTSRYEL--WGQGLTV 117  
 OY 120 SS 121  
 DB 118 SS 119

RESULT 13  
 D36005  
 IG heavy chain V region (M43) - human  
 C:Species: Homo sapiens (man)  
 C:Date: 21-Dec-1990 #sequence\_revision 21-Dec-1990 #text\_change 16-Dec-1998  
 C:Accession: D36005  
 R:Schroeder Jr., H.W.; Wang, J.Y.  
 Proc. Natl. Acad. Sci. U.S.A. 87, 6146-6150, 1990  
 A:Title: Preferential utilization of conserved immunoglobulin heavy chain variable gene  
 A:Reference number: A36005; PMID:90349571; PMID:2117273  
 A:Accession: D36005  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-119 <SCH>  
 A:Cross-references: GB:M34024  
 C:Genetics:  
 A:Gene: GDB:IGH@; IGHV1  
 A:Cross-references: GDB:118731; OMIM:146910  
 A:Map position: 14q32.33-14q32.33  
 C:Superfamily: immunoglobulin V region; immunoglobulin homology  
 C:Keywords: heterotetramer; immunoglobulin  
 F:15-98/Domain: immunoglobulin homology <IMM>

Query Match 53.4%; Score 365; DB 2; Length 119;  
 Best Local Similarity 63.9%; Pred. No. 2.8e-25;  
 Matches 78; Conservative 10; Mismatches 30; Indels 4; Gaps 2;

OY 1 QVQLSSEGGGLVQAGSLRLSCAASGSIHTMGYRQTPGERDVATIQ-DGGSITNY 59  
 DB 1 EVQLSSEGGGLVQPGSLRLSCAASGFTFSYMSWVRQAPGKLEWVAISGSGSTYY 60  
 OY 60 ADSVKGFTISRDNINLTNYVLOMNSLKPEPTAVYHCNADVRPYRTSRYEL--WGQGLTV 119  
 DB 61 ADSVKGFTISRDNINLTNYVLOMNSLKPEPTAVYHCNADVRPYRTSRYEL--WGQGLTV 117  
 OY 120 SS 121





GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2004, 07:29:35 ; Search time 11.6358 Seconds  
(without alignments)  
581.749 Million cell updates/sec

Title: US-09-805-290A-21

Perfect score: 683  
Sequence: 1 QVQLQESGGGLVQAGGSLRL.....WGQGLTVTVSSSEPTPKPKP 130

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Swissprot\_42.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	356	52.1	115	1	HV3F_HUMAN
2	342.5	50.1	122	1	HV3G_HUMAN
3	337	49.3	119	1	HV3I_HUMAN
4	328.5	48.1	126	1	HV3K_HUMAN
5	327	47.9	116	1	HV05_CARAU
6	326.5	47.8	119	1	HVAN_HUMAN
7	325.5	47.7	114	1	HV3B_HUMAN
8	324.5	47.5	119	1	HV3V_HUMAN
9	320.5	46.9	120	1	HV3U_HUMAN
10	318.5	46.6	114	1	HV01_CANFA
11	318.5	46.6	117	1	HV3C_HUMAN
12	318	46.6	115	1	HV3D_HUMAN
13	317.5	46.5	116	1	HV3T_HUMAN
14	315.5	46.2	136	1	HV16_MOUSE
15	314.5	46.0	122	1	HV3A_HUMAN
16	314	46.0	117	1	HV02_CANFA
17	314	46.0	121	1	HV3J_HUMAN
18	312	45.7	120	1	HV3E_HUMAN
19	311.5	45.6	119	1	HV3P_HUMAN
20	311	45.5	119	1	HV3L_HUMAN
21	308.5	45.2	98	1	HV57_MOUSE
22	308	45.1	115	1	HV33_MOUSE
23	307	44.9	97	1	HV56_MOUSE
24	306.5	44.9	119	1	HV40_MOUSE
25	306	44.8	115	1	HV32_MOUSE
26	305.5	44.7	117	1	HV30_HUMAN
27	305.5	44.7	119	1	HV37_MOUSE
28	304.5	44.6	117	1	HV54_MOUSE
29	304	44.5	113	1	HV30_MOUSE
30	303.5	44.4	122	1	HV20_MOUSE
31	302.5	44.3	117	1	HV53_MOUSE
32	302.5	44.3	117	1	HV55_MOUSE
33	302	44.2	123	1	HV24_MOUSE

34	301.5	44.1	122	1	HV3H_HUMAN
35	301	44.1	123	1	HV28_MOUSE
36	300.5	44.0	119	1	HV38_MOUSE
37	299	43.8	113	1	HV27_MOUSE
38	299	43.8	123	1	HV25_MOUSE
39	298	43.6	118	1	HV39_MOUSE
40	298	43.6	123	1	HV22_MOUSE
41	297.5	43.6	111	1	HV35_MOUSE
42	297.5	43.6	122	1	HV21_MOUSE
43	296	43.3	123	1	HV18_MOUSE
44	295	43.2	113	1	HV31_MOUSE
45	295	43.2	144	1	HV26_MOUSE

ALIGNMENTS

RESULT 1	ID	HV3F_HUMAN	STANDARD;	PRT;	115 AA.
AC	P01767;				
DT	21-JUL-1986 (Rel. 01, Created)				
DT	21-JUL-1986 (Rel. 01, Last sequence update)				
DT	10-OCT-2003 (Rel. 42, Last annotation update)				
DE	lg heavy chain V-II region BUT.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RP	SEQUENCE.				
RX	MEDLINE=78137069; PubMed=416441;				
RA	Torano A., Putnam F.W.;				
RT	"Complete amino acid sequence of the alpha 2 heavy chain of a human				
RT	lgA2 immunoglobulin of the A2m (2) allotype."				
RL	Proc. Natl. Acad. Sci. U.S.A. 75:966-969(1978).				
CC	-1- MISCELLANEOUS: THE SEQUENCE OF THE ALPHA-2, A2M(2) ALLOTYPE, C				
CC	REGION OF THIS MYELOMA PROTEIN IS ALSO GIVEN.				
CC	-1- SIMILARITY: Contains 1 immunoglobulin-like domain.				
DR	PIR; A02050; AZHUB.				
DR	HSSP; P01789; IMCP.				
DR	GO; GO:0005576; C:extracellular; NAS.				
DR	GO; GO:0003823; F:antigen binding; NAS.				
DR	GO; GO:0006955; P:immune response; NAS.				
DR	InterPro; IPR007110; Ig-like.				
DR	InterPro; IPR003596; Ig_v.				
DR	PIfam; PF00047; Ig; 1.				
DR	SMART; SM00406; IGV; 1.				
DR	PROSITE; PS50835; IG_LIKE; 1.				
KW	Immunoglobulin V region.				
FT	DOMAIN 1 111 IG-LIKE.				
FT	NON TER 115				
SQ	SEQUENCE 115 AA; 12379 MW; 208876A7DF52DC64;				
Query Match	52.1%; Score 356; DB 1; Length 115;				
Best Local Similarity	61.2%; Pred. No. 6.2e-29;				
Matches	74; Conservative 15; Mismatches 26; Indels 6; Gaps 1;				
QY	1 QVQLQESGGGLVQAGGSLRLISCAASGISGITHMGWYRQTPGTERDVAVTIQDGSITNYA 60				
DB	1 EVQLVETGGGLIQPGGSLRLISCAASGFTVSBHSMWROKPGKALZMWVSATYRGTTTYA 60				
QY	61 DSVKGRPTISRDNIITNYIQNNSLKPEDAVYHCNADYPRYTSSYELMGQGLTVTS 120				
DB	61 DSVKGRPTISRDNRITVYQNNSLKPEDAVYHCNADYPRYTSSYELMGQGLTVTS 114				
QY	121 S 121				
DB	115 S 115				
RESULT 2					
HV3G_HUMAN					

```

ID HV3G_HUMAN STANDARD; PRT; 122 AA.
AC P01768;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DE 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ig heavy chain V-III region CMM.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homindae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP MEDLINE=61013859; PubMed=6774332;
RA Lehman D.W., Putnam F.W.;
RT "Amino acid sequence of the variable region of a human mu chain:
RT location of a possible JH segment."
RL Proc. Natl. Acad. Sci. U.S.A. 77:3239-3243(1980).
CC -1- MISCELLANEOUS: THIS MU CHAIN WAS ISOLATED FROM THE PLASMA OF A
CC PATIENT WITH MACROGLOBULINEMIA.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.
DR PIR; A02051; M3H0AM.
DR HSSP; P01772; 2FB4.
DR GO; GO:0005576; C:extracellular; NAS.
DR GO; GO:0003823; F:antigen binding; NAS.
DR GO; GO:0006955; P:immune response; NAS.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003596; Ig_v.
DR Pfam; PF00047; Ig; 1.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS50835; IG_LIKE; 1.
KM Immunoglobulin V region; Pyroliidone carboxylic acid.
FT DOMAIN 1 112
FT MOD_RES 1 112
FT NON_TER 1 112
SQ SEQUENCE 122 AA; 13668 MW; A42D0F17D525F1C2 CRC64;

Query Match 50.1%; Score 342.5; DB 1; Length 122;
Best Local Similarity 59.0%; Pred. No. 1.5e-27;
Matches 72; Conservative 13; Mismatches 36; Indels 1; Gaps 1;

QY 1 QVQLQSGGGLVQAGGSLRLSCAASGSIHTMGYRTPGTEDVATIQ-DGGSINY 59
DB 1 QVELVSSGGGVVPGPSRLSLSCAASGFTFSNAMYRQPPKGLAWVAISYBGBKYY 60
QY 60 ADSVKGRFTISRDNINITYLQNNSLKPEDTAVYHCNADVRPYRTSRYLELWGQGLTVV 119
DB 61 ADSVKGRFTISRDNINITYLQNNSLKPEDTAVYHCNADVRPYRTSRYLELWGQGLTVV 120
QY 120 SS 121
DB 121 SS 122

RESULT 3
ID HV3I_HUMAN STANDARD; PRT; 119 AA.
AC P01770;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DE 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ig heavy chain V-III region NIE.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homindae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP MEDLINE=77070269; PubMed=826475;
RA Fenselau H., Hilschmann N.;
RT "The rule of antibody structure. The primary structure of a
RT monoclonal IgG1 immunoglobulin (myeloma protein NIE). III. The
RT chymotryptic peptides of the H-chain, alignment of the tryptic
RT peptides and discussion of the complete structure."

```

```

RL Hoppe-Seyler's Z. Physiol. Chem. 357:1571-1604(1976).
RN [2]
RP DISULFIDE BOND.
RX MEDLINE=77070267; PubMed=1002129;
RA Dreker L., Schwarz J., Reichel W., Hilschmann N.;
RT "Rule of antibody structure. The primary structure of a monoclonal
RT IgG1 immunoglobulin (myeloma protein NIE). I: Purification and
RT characterization of the protein, the L- and H-chains, the
RT cyanogen bromide cleavage products, and the disulfide bridges."
RL Hoppe-Seyler's Z. Physiol. Chem. 357:1515-1540(1976).
CC -1- MISCELLANEOUS: THIS CHAIN WAS ISOLATED FROM AN IGG1 MYELOMA
CC PROTEIN.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.
DR PIR; A91689; G1H0NT.
DR HSSP; P01772; 2FB4.
DR GO; GO:0005576; C:extracellular; NAS.
DR GO; GO:0003823; F:antigen binding; NAS.
DR GO; GO:0006955; P:immune response; NAS.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003596; Ig_v.
DR Pfam; PF00047; Ig; 1.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS50835; IG_LIKE; 1.
KM Immunoglobulin V region; Pyroliidone carboxylic acid.
FT DOMAIN 1 112
FT MOD_RES 1 112
FT NON_TER 1 112
SQ SEQUENCE 119 AA; 13242 MW; C96933A6E55E165B CRC64;

Query Match 49.3%; Score 337; DB 1; Length 119;
Best Local Similarity 59.8%; Pred. No. 5.2e-27;
Matches 73; Conservative 16; Mismatches 29; Indels 4; Gaps 3;

QY 1 QVQLQSGGGLVQAGGSLRLSCAASGSIHTMGYRTPGTEDVATIQ-DGGSINY 59
DB 1 QVELVSSGGGVVPGPSRLSLSCAASGFTFSRYTHVWRAPGGLAWVAISYBGBKYY 60
QY 60 ADSVKGRFTISRDNINITYLQNNSLKPEDTAVYHCNADVRPYRTSRYLELWGQGLTVV 119
DB 61 ADSVKGRFTISRDNINITYLQNNSLKPEDTAVYHCNADVRPYRTSRYLELWGQGLTVV 117
QY 120 SS 121
DB 118 SS 119

RESULT 4
ID HV3K_HUMAN STANDARD; PRT; 126 AA.
AC P01772;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DE 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ig heavy chain V-III region KOL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homindae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP MEDLINE=83289131; PubMed=6884994;
RA Schmidt W.E., Jung H.-D., Palm M., Hilschmann N.;
RT "Three-dimensional structure determination of antibodies. Primary
RT structure of crystallized monoclonal immunoglobulin IgG1 KOL. I."
RL Hoppe-Seyler's Z. Physiol. Chem. 364:713-747(1983).
RN [2]
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).
RX MEDLINE=81072295; PubMed=7441755;
RA Marquart M., Dessenhofer J., Huber R., Palm M.;
RT "Crystallographic refinement and atomic models of the intact
RT immunoglobulin molecule KOL and its antigen-binding fragment at 3.0 A
RT and 1.0-A resolution."

```



RL J. Mol. Biol. 141:369-391(1980).  
CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.  
DR PIR; A02055; G1HOKL.  
DR PDB; 2PB4; 12-JUL-89.  
DR PDB; 2IG2; 12-JUL-89.  
DR GO; GO:0005576; C:extracellular; NAS.  
DR GO; GO:0003823; F:antigen binding; NAS.  
DR GO; GO:0006955; P:immune response; NAS.  
DR InterPro; IPR007110; IG-like.  
DR InterPro; IPR003596; Ig\_v.  
DR Pfam; PF00047; Ig; 1.  
DR SMART; SM00406; IGV; 1.  
DR PROSITE; PS50835; IG\_LIKE; 1.  
KM Immunoglobulin V region; 3D-structure; Pyrrolidone carboxylic acid.  
FT DOMAIN 1 112 IG-LIKE  
FT MOD\_RES 1 1 PYRROLIDONE CARBOXYLIC ACID.  
FT DISULFID 22 96  
FT STRAND 105 110  
FT STRAND 3 7  
FT STRAND 11 12  
FT TURN 14 15  
FT STRAND 18 25  
FT HELIX 29 31  
FT STRAND 34 39  
FT TURN 41 42  
FT STRAND 45 51  
FT TURN 53 54  
FT STRAND 58 60  
FT HELIX 62 64  
FT STRAND 65 65  
FT TURN 66 67  
FT STRAND 68 73  
FT TURN 74 77  
FT STRAND 78 83  
FT HELIX 88 90  
FT STRAND 92 99  
FT STRAND 106 106  
FT TURN 107 108  
FT STRAND 109 109  
FT STRAND 113 116  
FT STRAND 120 124  
FT NON\_TER 126 126  
SQ SEQUENCE 126 AA; 13718 MW; E4D71B52B16F8776 CRC64;  
Query Match 48.1%; Score 328.5; DB 1; Length 126;  
Best Local Similarity 56.4%; Pred. No. 4e-26;  
Matches 75; Conservative 11; Mismatches 28; Indels 19; Gaps 3;  
QY 1 QVQLQESGGGLVQAGGSLRLSCAASGSIIGTHTMGWRQTPGTERTDVAATIQDGGSTNY 59  
DB 1 QVQLVESGGGLVQPGSRSLRLSCSSGFLFSSYAMVWRQAPGKGLWVAITWDSQDHY 60  
QY 60 ADYKGRFTISRDNILNTVYLQWNSLKEPEDTAVYHCNAD-----VRPYRTSRYL 108  
DB 61 ADYKGRFTISRDNILNTVYLQWNSLKEPEDTAVYHCNAD-----VRPYRTSRYL 113  
QY 109 ELWGQGLIVTYS 121  
DB 114 DYGQGTPTVVS 126  
RESULT 5  
HV05 CARAU STANDARD; PRT; 116 AA.  
AC P19181;  
DT 01-NOV-1990 (Rel. 16, Created)  
DT 01-NOV-1990 (Rel. 16, Last sequence update)  
DE 15-JUL-1999 (Rel. 38, Last annotation update)  
DE Ig heavy chain V region 5A precursor.  
OS Carassius auratus (Goldfish).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Osteichthys; Cypriniformes;  
OC Cyprinidae; Carassius.

OX NCBI\_TaxID=7957;  
RN [1]  
RE SEQUENCE FROM N.A.  
RX MEDLINE=88144476; PubMed=3125551;  
RA Wilson M.R., Middleton D., Watt G.W.;  
RT "Immunoglobulin heavy chain variable region gene evolution: structure and family relationships of two genes and a pseudogene in a teleost fish."  
RU Proc. Natl. Acad. Sci. U.S.A. 85:1566-1570(1988).  
DR HSSP; P01772; 2PB4.  
DR InterPro; IPR007110; IG-like.  
DR InterPro; IPR003596; Ig\_v.  
DR Pfam; PF00047; Ig; 1.  
DR SMART; SM00406; IGV; 1.  
DR PROSITE; PS50835; IG\_LIKE; 1.  
KM Immunoglobulin V region; Signal.  
FT SIGNAL 1 19  
FT CHAIN 20 116 IG HEAVY CHAIN V REGION 5A.  
FT DOMAIN 20 49 FRAMEWORK-1.  
FT DOMAIN 50 54 COMPLEMENTARITY-DETERMINING-1.  
FT DOMAIN 55 68 FRAMEWORK-2.  
FT DOMAIN 69 84 COMPLEMENTARITY-DETERMINING-2.  
FT DOMAIN 85 116 FRAMEWORK-3.  
FT DISULFID 41 114 BY SIMILARITY.  
FT NON\_TER 116 116  
SQ SEQUENCE 116 AA; 12808 MW; 9C2279E2DF199B12 CRC64;  
Query Match 47.9%; Score 327; DB 1; Length 116;  
Best Local Similarity 69.5%; Pred. No. 5.1e-26;  
Matches 66; Conservative 8; Mismatches 21; Indels 0; Gaps 0;  
QY 1 QVQLQESGGGLVQAGGSLRLSCAASGSIIGTHTMGWRQTPGTERTDVAATIQDGGSTNYA 60  
DB 20 EVQLVESGGGLVQPGSRSLRLSCSSGFLFSSYAMVWRQAPGKGLWVAITWDSQDHY 79  
QY 61 DSYKGRFTISRDNILNTVYLQWNSLKEPEDTAVYHC 95  
DB 80 DSYKGRFTISRDNILNTVYLQWNSLKEPEDTAVYHC 114  
RESULT 6  
HV03 HUMAN STANDARD; PRT; 119 AA.  
AC P01775;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Ig heavy chain V-II region LAY.  
OS Homo sapiens (human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RE SEQUENCE.  
RX MEDLINE=75046755; PubMed=4139708;  
RA Capra J.D., Kehoe J.M.;  
RT "Structure of antibodies with shared idiotype: the complete sequence of the heavy chain variable regions of two immunoglobulin K anti-gamma globulins."  
RU Proc. Natl. Acad. Sci. U.S.A. 71:4032-4036(1974).  
CC -1- MISCELLANEOUS: THIS CHAIN WAS ISOLATED FROM IGM WITH ANTI-GAMMA GLOBULIN ACTIVITY.  
CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.  
DR PIR; A02058; M3HUX.  
DR HSSP; P01772; 2PB4.  
DR GO; GO:0005576; C:extracellular; NAS.  
DR GO; GO:0003823; F:antigen binding; NAS.  
DR GO; GO:0006955; P:immune response; NAS.  
DR InterPro; IPR007110; IG-like.  
DR InterPro; IPR003596; Ig\_v.  
DR Pfam; PF00047; Ig; 1.  
DR SMART; SM00406; IGV; 1.  
DR PROSITE; PS50835; IG\_LIKE; 1.

KM Immunoglobulin V region.  
 FT DOMAIN 1 112 IG-LIKE.  
 SQ NON TER 119 119  
 SEQUENCE 119 AA; 12858 MW; D6338098794DCFE5 CRC64;

Query Match  
 Best Local Similarity 47.8%; Score 326.5; DB 1; Length 119;  
 Matches 67; Conservative 14; Mismatches 36; Indels 1; Gaps 1;

QY 2 VOLOESGGGLVQAGSLRLSCAASGTSIGSIHTMGVYRQTPGERPVVA-TIODGGSITYA 60  
 DB 2 VOLESGLVQPGGSLRLSCAASGFTTSASMSVVRQAPGKGLVWAKTENGNDKRYA 61

QY 61 DSVKGRFTISRDNINITYLNQNSLKPEDTAVYHGNADVRPRYSRYLELWGQGLTVT 118  
 DB 62 DSVNKRFTISRNDKNTLYLNQNGLQAZVSAIYCARAGPYVSPFPFHWGQGLTVT 119

RESULT 7  
 HV3B HUMAN STANDARD; PRT; 114 AA.  
 ID HV3B HUMAN  
 AC P01763;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Ig heavy chain V-II region WEA.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 NC NCBI\_TaxID=9606;  
 RN [1]  
 RP MEDLINE=83273707; PubMed=6410398;  
 RA Gont F., Frangione B.;  
 RT "Amino acid sequence of the Fv region of a human monoclonal Igm  
 (protein WEA) with antibody activity against 3,4-pyruvylated  
 galactose in Klebsiella polysaccharides K30 and K33."  
 RL Proc. Natl. Acad. Sci. U.S.A. 80:4837-4841(1983).  
 CC -1- MISCELLANEOUS. THIS CHAIN WAS OBTAINED FROM A MONOCLONAL ANTIBODY  
 AGAINST 3,4-PYRUVYLATED GALACTOSE AND ISOLATED FROM A PATIENT WITH  
 WALSLEYFROM'S MACROGLOBULINEMIA.  
 CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.  
 DR PIR; A02046; M3HWE.  
 DR HSSP; P01772; 2F84.  
 DR GO; GO:0005576; C:extracellular; NAS.  
 DR GO; GO:0003823; F:antigen binding; NAS.  
 DR GO; GO:0006955; P:immune response; NAS.  
 DR InterPro; IPR007110; IG\_V.  
 DR InterPro; IPR003596; IG\_V.  
 DR Pfam; PF00047; IGV; 1.  
 DR SMART; SM00406; IGV; 1.  
 DR PROSITE; PS50835; IG\_LIKE; 1.  
 KM Immunoglobulin V region; Pyroliidone carboxylic acid.  
 FT DOMAIN 1 112 IG-LIKE.  
 FT MOD RES 1 114 PYROLIDONE CARBOXYLIC ACID.  
 FT NON TER 114 114  
 SQ SEQUENCE 114 AA; 12256 MW; D88294FB418A07B7 CRC64;

Query Match  
 Best Local Similarity 47.7%; Score 325.5; DB 1; Length 114;  
 Matches 71; Conservative 15; Mismatches 27; Indels 9; Gaps 2;

QY 1 QVOLOESGGGLVQAGSLRLSCAASGTSIGSIHTMGVYRQTPGERPVVA-TIODGGSITYA 59  
 DB 1 QVQLVDSGGGLVQPGGSLRLSCAASGFTTSASMSVVRQAPGKGLVWAKTENGNDKRYA 60

QY 60 ADSVKGRFTISRDNINITYLNQNSLKPEDTAVYHGNADVRPRYSRYLELWGQGLTVT 119  
 DB 61 ADSVKGRFTISRDNINITYLNQNSLKPEDTAVYHGNADVRPRYSRYLELWGQGLTVT 112

QY 120 SS 121  
 DB 113 SS 114

RESULT 8  
 HV3M HUMAN STANDARD; PRT; 119 AA.  
 ID HV3M HUMAN  
 AC P01774;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Ig heavy chain V-II region POW.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 NC NCBI\_TaxID=9606;  
 RN [1]  
 RP MEDLINE=75046755; PubMed=4139708;  
 RA Capra J.D., Kehoe J.M.;  
 RT "Structure of antibodies with shared idiotypic: the complete sequence  
 of the heavy chain variable regions of two immunoglobulin M  
 anti-gamma globulins."  
 RL Proc. Natl. Acad. Sci. U.S.A. 71:4032-4036(1974).  
 CC -1- MISCELLANEOUS. THIS CHAIN WAS ISOLATED FROM IGM WITH ANTI-GAMMA  
 GLOBULIN ACTIVITY.  
 CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.  
 DR PIR; A02057; M3HUPM.  
 DR HSSP; P01772; 2F84.  
 DR GO; GO:0005576; C:extracellular; NAS.  
 DR GO; GO:0003823; F:antigen binding; NAS.  
 DR GO; GO:0006955; P:immune response; NAS.  
 DR InterPro; IPR007110; IG\_V.  
 DR InterPro; IPR003596; IG\_V.  
 DR Pfam; PF00047; IGV; 1.  
 DR SMART; SM00406; IGV; 1.  
 DR PROSITE; PS50835; IG\_LIKE; 1.  
 KM Immunoglobulin V region.  
 FT DOMAIN 1 112 IG-LIKE.  
 FT VARIANT 54 54 N -> D (PROBABLY DUE TO DEAMINATION  
 DURING ISOLATION).  
 FT FT /FTID=VAR\_003966.  
 FT NON TER 119 119  
 SQ SEQUENCE 119 AA; 12953 MW; 2E018AF4DCBB2610 CRC64;

Query Match  
 Best Local Similarity 47.5%; Score 324.5; DB 1; Length 119;  
 Matches 67; Conservative 16; Mismatches 35; Indels 1; Gaps 1;

QY 1 QVOLOESGGGLVQAGSLRLSCAASGTSIGSIHTMGVYRQTPGERPVVA-TIODGGSITYA 59  
 DB 1 EVQLDSGGGLVQPGGSLRLSCAASGFTSSASMSVVRQAPGKGLVWAKTENGNDKRYA 60

QY 60 ADSVKGRFTISRDNINITYLNQNSLKPEDTAVYHGNADVRPRYSRYLELWGQGLTVT 118  
 DB 61 ADSVKGRFTISRNDKNTLYLNQNGLQAZVSAIYCARAGPYVSPFPFHWGQGLTVT 119

RESULT 9  
 HV3U HUMAN STANDARD; PRT; 120 AA.  
 ID HV3U HUMAN  
 AC P01782;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Ig heavy chain V-II region DOB.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 NC NCBI\_TaxID=9606;  
 RN [1]  
 RP MEDLINE=80020921; PubMed=114209;  
 RA Steiner L.A., Garcia Pardo A., Margolies M.N.;  
 RT "Amino acid sequence of the heavy-chain variable region of the

```

RT crystalizable human myeloma protein Dob.;
RL Biochemistry 18:4068-4080(1979).
RN [2]
RP CRYSTALLIZATION.
RX MEDLINE=80020920; PubMed=114208;
RA Steiner L.A.; Lopes A.D.;
RT "write crystalizable human myeloma protein Dob has a hinge-region
deletion."
RL Biochemistry 18:4054-4067(1979).
CC -1- MISCELLANEOUS: THIS GAMMA-1 MYELOMA PROTEIN HAS A DELETION IN THE
HINGE REGION. THERE ARE NO LIGHT-HEAVY OR INTER-HEAVY CHAIN
DISULFIDE BONDS.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.
DR PIR; A20431; GIHDB.
DR HSSP; P01772; 2PB4.
DR GO; GO:0005576; C:extracellular; NAS.
DR GO; GO:0003823; F:antigen binding; NAS.
DR GO; GO:0006955; P:immune response; NAS.
DR InterPro; IPR007110; IG_1-like.
DR InterPro; IPR003596; IG_V.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS50835; IG-LIKE; 1.
KW Immunoglobulin V region.
FT DOMAIN 1 112 IG-LIKE.
FT NON_TER 120 120
SQ SEQUENCE 120 AA; 13440 MW; 880DD307C4B2627 CRC64;

Query Match 46.9%; Score 320.5; DB 1; Length 120;
Best Local Similarity 58.5%; Pred. No. 2.4e-25;
Matches 72; Conservative 12; Mismatches 34; Indels 5; Gaps 3;

QY 1 QVQLQESGGGLVQAGGSLRLSCAASGSIHTMGWVRQTPGTERDVAIIQ-DGGSTNY 59
DB 1 EVQLVESGGDLVQPGSRSLRLSCAASGFNFHMYNHWTLRQCGKGPKEWVSITTMNGGSVLY 60
QY 60 ADVSKRFTISRDNIINTVYLQNSLKPEDTAVYHCNADVAPY-RTSYLELMQGGTLTV 118
DB 61 ADVSKRFTISRDNAKQTLTYQLNLRPEDTAVFYC--AKGYLWNGNWPMSQGGTLTV 117
QY 119 VSS 121
DB 118 VSS 120
DB 118 VSS 120

RESULT 10
HVO1_CANFA STANDARD; PRT; 114 AA.
ID HVO1_CANFA STANDARD; PRT; 114 AA.
AC P01784;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE 1g heavy chain V region GDM.
OS Carls familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_Taxid=9615;
RN [1]
RP SEQUENCE.
RX MEDLINE=77242268; PubMed=407924;
RA Wasserman R.L.; Capra J.D.;
RT "Primary structure of the variable regions of two canine
immunoglobulin heavy chains."
RL Biochemistry 16:3160-3168(1977).
CC -1- MISCELLANEOUS: THIS CHAIN WAS ISOLATED FROM A MYELOMA PROTEIN.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.
DR PIR; A20267; AVDCGM.
DR HSSP; P01772; 2PB4.
DR InterPro; IPR007110; IG_1-like.
DR InterPro; IPR003596; IG_V.
DR Pfam; PF00047; IG; 1.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS50835; IG-LIKE; 1.

```

```

KM Immunoglobulin V region.
FT DOMAIN 1 112 IG-LIKE.
FT NON_TER 114 114
SQ SEQUENCE 114 AA; 12430 MW; B1D4745D2C4E13C4 CRC64;

Query Match 46.6%; Score 318.5; DB 1; Length 114;
Best Local Similarity 57.4%; Pred. No. 3.6e-25;
Matches 70; Conservative 13; Mismatches 30; Indels 9; Gaps 3;

QY 1 QVQLQESGGGLVQAGGSLRLSCAASGSIHTMGWVRQTPGTERDVAIIQ-DGSTNY 59
DB 1 EVQLVESGGDLVQPGSRSLRLSCAASGFNFHMYNHWTLRQCGKGPKEWVSITTMNGGSVLY 60
QY 60 ADVSKRFTISRDNIINTVYLQNSLKPEDTAVYHCNADVAPY-RTSYLELMQGGTLTV 119
DB 61 ADVSKRFTISRDNAKQTLTYQLNLRPEDTAVFYC--AKGYLWNGNWPMSQGGTLTV 117
QY 120 SS 121
DB 113 SS 114

RESULT 11
HVO3C_HUMAN STANDARD; PRT; 117 AA.
ID HVO3C_HUMAN STANDARD; PRT; 117 AA.
AC P01764;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE 1g heavy chain V-III region VH26 precursor.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=81101090; PubMed=6450418;
RA Macthysens G.; Rabbitts T.H.;
RT "Structure and multiplicity of genes for the human immunoglobulin
heavy chain variable region."
RL Proc. Natl. Acad. Sci. U.S.A. 77:6561-6565(1980).
CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
DR EMBL; J00236; AAA53516.1; -
DR EMBL; M35415; AAA58735.1; -
DR PIR; A02047; H3HU26.
DR PDB; 1HOU; 23-DEC-99.
DR Genew; HGNC:5545; IGHV@.
DR GO; GO:0005576; C:extracellular; NAS.
DR GO; GO:0003823; F:antigen binding; NAS.
DR GO; GO:0006955; P:immune response; NAS.
DR InterPro; IPR007110; IG_1-like.
DR InterPro; IPR003596; IG_V.
DR Pfam; PF00047; IG; 1.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS50835; IG-LIKE; 1.
KW Immunoglobulin V region; Signal; 3D-structure.
FT SIGNAL 1 19
FT CHAIN 20 117 IG HEAVY CHAIN V-III REGION VH26.
FT DOMAIN 20 >117 IG-LIKE.
FT NON_TER 117 117
SQ SEQUENCE 117 AA; 12582 MW; E826733F1A3CB0F1 CRC64;

Query Match 46.6%; Score 318.5; DB 1; Length 117;
Best Local Similarity 68.8%; Pred. No. 3.7e-25;

```

Matches 66; Conservative 7; Mismatches 22; Indels 1; Gaps 1;

QY 1 QVQLQSSGGGLVQAGSLRLSCAASGSIISGTHMTGWYRQTGTERDVAATIQ-DGGSTNY 59  
 DB 20 EVQLLESGGGLVQPGGSLRLSCAASGFTFSSYAMSWVQAPEKGLWVAISGGGGSTYY 79  
 QY 60 ADVKGRFTISRDNILNTVYLQMSLKPEDTAVVHC 95  
 DB 80 GDSVKGRTISRDNISNTVYLQMSLRAEDTAVVYC 115

RESULT 12

HVAD HUMAN

ID HV3J HUMAN STANDARD; PRT; 115 AA.

AC P01765;

DT 21-JUL-1986 (Rel. 01, Created)

DT 21-JUL-1986 (Rel. 01, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Ig heavy chain V-II region TIL.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

OC NCBI\_TaxID=9606;

OX NCBI\_TaxID=9606;

RX NCBI\_TaxID=9606;

RE MEDLINE=78005528; PubMed=409716;

RA Wang A.-C., Wang I.Y., Fudenberg H.H.;

RT "Immunoglobulin structure and genetics. Identity between variable

regions of a mu and a gamma chain.";

J Biol. Chem. 253:7192-7199 (1977).

CC -1- MISCELLANEOUS: THE SEQUENCES OF THE V REGIONS OF THE HEAVY CHAINS

OF IGM AND IGG2 ISOLATED FROM A SINGLE PATIENT WITH BICLONAL

GAMMOPATHY ARE IDENTICAL. THEIR LIGHT CHAINS ARE APPARENTLY ALSO

IDENTICAL.

CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.

DR PIR; A02048; H3HUTL.

DR HSSP; P01772; 2FB4.

DR GO; GO:0005576; C:extracellular; NAS.

DR GO; GO:0003823; P:antigen binding; NAS.

DR GO; GO:0006955; P:immune response; NAS.

DR InterPro; IPR007110; Ig-like.

DR Pfam; PF00047; IgV\_1.

DR SMART; SM00406; IGV\_1.

DR PROSITE; PSS0835; IG LIKE; 1.

KM Immunoglobulin V region.

FT DOMAIN 1 108 IG-LIKE.

FT NON\_TER 115 115

SC SEQUENCE 115 AA; 12356 MW; 4DCC67D179F62326 CRC64;

Query Match 46.6%; Score 318; DB 1; Length 115;

Best Local Similarity 58.5%; Pred. No. 4.1e-25;

Matches 72; Conservative 11; Mismatches 30; Indels 10; Gaps 4;

QY 1 QVQLQSSGGGLVQAGSLRLSCAASGSIISGTHMTGWYRQTGTERDVAATIQ-DGGSTNY 59

DB 1 EVQLLESGGGLVQPGGSLRLSCAASGFTFSSYAMSWVQAPEKGLWVAISGGGGSTYY 60

QY 60 ADVKGRFTISRDNILNTVYLQMSLKPEDTAVVHC-NADVPYRTSYELMGGGLTVT 118

DB 61 ABSVKGRTISRDNISNTVYLQMSLRAEDTAVVYC---YFVYWGGLTVT 112

QY 119 VSS 121

DB 113 VSS 115

RESULT 13

HV3J HUMAN

ID HV3J HUMAN STANDARD; PRT; 116 AA.

AC P01761;

DT 21-JUL-1986 (Rel. 01, Created)

DT 21-JUL-1986 (Rel. 01, Last sequence update)

DE Ig heavy chain V-II region TIL.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

OC NCBI\_TaxID=9606;

OX NCBI\_TaxID=9606;

RX NCBI\_TaxID=9606;

RE MEDLINE=75059123; PubMed=4803843;

RA Watanabe S., Barnikol H.U., Horn J., Bertram J., Hilschmann N.;

RT "The primary structure of a monoclonal IgM-immunoglobulin

(macroglobulin Gal.) II: the amino acid sequence of the H-chain (mu-

type), subgroup H III: Architecture of the complete IgM-molecule.";

RL Hoppe-Seyler's Z. Physiol. Chem. 354:1505-1509 (1973).

DE 10-OCT-2003 (Rel. 42, Last annotation update)

DE Ig heavy chain V-II region GAL.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

OC NCBI\_TaxID=9606;

OX NCBI\_TaxID=9606;

RX NCBI\_TaxID=9606;

RE MEDLINE=75059123; PubMed=4803843;

RA Watanabe S., Barnikol H.U., Horn J., Bertram J., Hilschmann N.;

RT "The primary structure of a monoclonal IgM-immunoglobulin

(macroglobulin Gal.) II: the amino acid sequence of the H-chain (mu-

type), subgroup H III: Architecture of the complete IgM-molecule.";

RL Hoppe-Seyler's Z. Physiol. Chem. 354:1505-1509 (1973).

FT DOMAIN 1 112 IG-LIKE.

FT NON\_TER 116 116

SC SEQUENCE 116 AA; 12730 MW; 2C67CA9AAAAA1282 CRC64;

Query Match 46.5%; Score 317.5; DB 1; Length 116;

Best Local Similarity 58.2%; Pred. No. 4.6e-25;

Matches 71; Conservative 12; Mismatches 32; Indels 7; Gaps 3;

QY 1 QVQLQSSGGGLVQAGSLRLSCAASGSIISGTHMTGWYRQTGTERDVAATIQ-DGGSTNY 59

DB 1 EVQLLESGGGLVQPGGSLRLSCAASGFTFSSYAMSWVQAPEKGLWVAISGGGGSTYY 60

QY 60 ADVKGRFTISRDNILNTVYLQMSLKPEDTAVVHC-NADVPYRTSYELMGGGLTVT 119

DB 61 VDSVKGRTISRDNISNTVYLQMSLRAEDTAVVYC---ARGWGGDY---WGQGLTVT 114

QY 120 SS 121

DB 115 ST 116

RESULT 14

HV16 MOUSE

ID HV16 MOUSE STANDARD; PRT; 136 AA.

AC P01783;

DT 21-JUL-1986 (Rel. 01, Created)

DT 21-JUL-1986 (Rel. 01, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Ig heavy chain V region MOPC 21 precursor (Fragment).

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OC NCBI\_TaxID=10090;

OX NCBI\_TaxID=10090;

RX NCBI\_TaxID=10090;

RE MEDLINE=81234546; PubMed=6788376;

RA Bothwell A.L.N., Paskind M., Rehn M., Imanishi-Kari T., Rajewsky K.,

Baltimore D.;

RT "Heavy chain variable region contribution to the NPb family of

antibodies: somatic mutation evident in a gamma 2a variable region.";

Cell 24:625-637 (1981).

```

RN [2]
RX SEQUENCE OF 17-136.
RX MEDLINE=77100368; PubMed=401950;
RA Adeyugo K., Miletin C., Secher D.S.;
RA "Molecular analysis of spontaneous somatic mutants.";
RL Nature 265:299-304(1977).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.ebi.ac.uk/announcements/
CC or send an email to license@ebi.ac.uk).
CC -----
CC EMBL: J00522; A015290.1; -.
DR PIR: E08089; G1MS21.
DR PDB: 1IGC; 03-JUN-95.
DR InterPro: IPR007110; IG-like.
DR InterPro: IPR003596; IG_v.
DR Pfam: PF00047; Ig_1.
DR SMART: SM00406; Ig; 1.
DR PROSITE: PS50835; IG-LIKE; 1.
KW Immunoglobulin V region; Signal; 3D-structure.
FT SIGNAL 1
FT CHAIN 1
FT DOMAIN 115 136 IG HEAVY CHAIN V REGION MOPC 21.
FT DOMAIN 120 136 D SEGMENT.
FT DISULFID 38 112 JH4 SEGMENT.
FT CONFLICT 75 78 HYAD -> DYAR (IN REF. 2).
FT CONFLICT 89 90 DN -> ND (IN REF. 2).
FT CONFLICT 115 115 W -> H (IN REF. 2).
FT CONFLICT 120 120 Y -> W (IN REF. 2).
FT NON_TER 136
SQ SEQUENCE 136 AA; 15071 MW; 2276A98DBDF7016 CRC64;

Query Match 46.2%; Score 315.5; DB 1; Length 136;
Best Local Similarity 57.6%; Pred. No. 8.9e-25;
Matches 72; Conservative 11; Mismatches 31; Indels 11; Gaps 3;

QY 2 VQLOESGGGLVQAGGSLRLSCAAGSISGSIHTMGYRQTPGERDVAATIDGGST-NYA 60
DB 18 VQVESGGGLVQPGGSRLLSCAAGSFPSFGMHWVRQAPKGLRWVAISSGSSLTAYA 77
QY 61 DSVKGRFTISRDNILNTYVQLQNSLKPEDTAVYHCNADVRPRTSRV---LELWGQGL 116
DB 78 DTVKGRFTISRDNILNTYVQLQNSLKPEDTAVYHCNADVRPRTSRV---LELWGQGL 131
QY 117 VTWSS 121
DB 132 VTWSS 136

RESULT 15
HV3A HUMAN STANDARD; PRT; 122 AA.
AC P01762;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ig heavy chain V-II region TRO.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE (MYELOMA PROTEIN TRO).
RX MEDLINE=76023781; PubMed=809331;
RA Kretzlin H., Altevogt P., Rudan E., Kortt A., Starescuk K.,
RA Hilschmann N.;
RT "The primary structure of a monoclonal IgA-immunoglobulin (IgA Tro.);
RT I. The amino acid sequence of the H-chain, alpha-type, subgroup III;

```

```

RT structure of the complete IgA-molecule.";
RL Hoppe-Seyler's Z. Physiol. Chem. 356:1337-1342(1975).
CC -1- MISCELLANEOUS: THE SEQUENCE OF THE C REGION IS ALSO GIVEN.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.
DR PIR: A02045; A1HUTR.
DR HSSP: P01772; 2FBA.
DR GO: GO:0005576; Cellular; NAS.
DR GO: GO:0003823; Antigen binding; NAS.
DR GO: GO:0006955; Immune response; NAS.
DR InterPro: IPR007110; IG-like.
DR InterPro: IPR003596; IG_v.
DR Pfam: PF00047; Ig_1.
DR SMART: SM00406; Ig; 1.
DR PROSITE: PS50835; IG-LIKE; 1.
KW Immunoglobulin V region; Pyroglutamate carboxylic acid.
FT MOD_RES 1 108 PYROGLUTAMATE CARBOXYLIC ACID.
FT DOMAIN 1
FT NON_TER 1
SQ SEQUENCE 122 AA; 13472 MW; 2E21A11DA04D8CF9 CRC64;

Query Match 46.0%; Score 314.5; DB 1; Length 122;
Best Local Similarity 55.7%; Pred. No. 9.8e-25;
Matches 68; Conservative 17; Mismatches 36; Indels 1; Gaps 1;

QY 1 QVQLOESGGGLVQAGGSLRLSCAAGSISGSIHTMGYRQTPGERDVAATIDGGST-NY 59
DB 1 QVQVQSGGGLVQPGGSRLLSCAAGSFPSFGMHWVRQAPKGLRWVAISSGSSLTYY 60
QY 60 ADSVKGRTISRDNILNTYVQLQNSLKPEDTAVYHCNADVRPRTSRVLELWGQGLTVV 119
DB 61 ADSVKGRTISRDNILNTYVQLQNSLKPEDTAVYHCNADVRPRTSRVLELWGQGLTVV 120
QY 120 SS 121
DB 121 SS 122

Search completed: October 5, 2004, 08:01:55
Job time: 12.6358 secs

```

This Page Blank (uspofo)

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2004, 07:34:11 ; Search time 68.8117 Seconds  
(without alignments)  
596.081 Million cell updates/sec

Title: US-09-805-290A-21

Perfect score: 683  
Sequence: 1 QVQLQESGGGLVQAGGSLRL.....WGQGLTVTVSSSEPTKPKPP 130

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: SP\_archaea:\*  
2: SP\_bacteria:\*  
3: SP\_fungi:\*  
4: SP\_human:\*  
5: SP\_invertebrate:\*  
6: SP\_mammal:\*  
7: SP\_mhc:\*  
8: SP\_organelle:\*  
9: SP\_phage:\*  
10: SP\_plant:\*  
11: SP\_todent:\*  
12: SP\_virus:\*  
13: SP\_vertebrate:\*  
14: SP\_unclassified:\*  
15: SP\_virus:\*  
16: SP\_bacterioplasmid:\*  
17: SP\_archaeoplasmid:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	374	54.8	597	4 Q96B89	Q96B89 homo sapien
2	365.5	53.5	119	4 Q9UL72	Q9UL72 homo sapien
3	361.5	52.9	487	11 Q99KA4	Q99KA4 mus musculu
4	359.5	52.6	499	4 Q8N5K4	Q8N5K4 mus musculu
5	358	52.4	113	4 Q9UL90	Q9UL90 mus musculu
6	348.5	51.0	613	4 Q8WUK1	Q8WUK1 mus musculu
7	346	50.7	121	4 Q9UL71	Q9UL71 mus musculu
8	346	50.7	486	11 Q91Z07	Q91Z07 mus musculu
9	342	50.1	480	11 Q91X21	Q91X21 mus musculu
10	340	49.8	119	11 Q92027	Q92027 mus musculu
11	339	49.6	131	4 Q9UL88	Q9UL88 mus musculu
12	339	49.6	147	4 Q9Y509	Q9Y509 mus musculu
13	337	49.3	479	11 Q91W5	Q91W5 mus musculu
14	337	49.3	573	4 Q8WU38	Q8WU38 mus musculu
15	336.5	49.3	118	4 Q9UL91	Q9UL91 mus musculu
16	336.5	49.3	493	4 Q8N5C6	Q8N5C6 mus musculu

17	335.5	49.1	112	4 Q9HCC1	Q9HCC1 mus musculu
18	335.5	49.1	122	4 Q9UL84	Q9UL84 mus musculu
19	335	49.0	116	4 Q9UL93	Q9UL93 mus musculu
20	334	48.9	494	4 Q96K68	Q96K68 mus musculu
21	324.5	47.5	473	11 Q91Z05	Q91Z05 mus musculu
22	322	47.1	487	11 Q80217	Q80217 mus musculu
23	321.5	47.1	469	11 Q8R3V9	Q8R3V9 mus musculu
24	320.5	46.9	521	4 Q8N4Y9	Q8N4Y9 mus musculu
25	318.5	46.6	437	11 Q91A4	Q91A4 mus musculu
26	313.5	45.9	95	4 Q9ULB6	Q9ULB6 mus musculu
27	308.5	45.2	228	11 Q9QIF0	Q9QIF0 mus musculu
28	303.5	44.4	470	4 Q7Z5M1	Q7Z5M1 mus musculu
29	300.5	44.0	124	6 Q9N0M6	Q9N0M6 mus musculu
30	298.5	43.7	124	6 Q9N0M4	Q9N0M4 mus musculu
31	298.5	43.7	484	11 Q8VEA0	Q8VEA0 mus musculu
32	279.5	40.9	482	4 Q7Z351	Q7Z351 mus musculu
33	276.5	40.5	479	11 Q7TMC4	Q7TMC4 mus musculu
34	265.5	38.9	124	4 Q9UL92	Q9UL92 mus musculu
35	262	38.4	104	4 Q9UL87	Q9UL87 mus musculu
36	261.5	38.3	482	11 Q91X92	Q91X92 mus musculu
37	261	38.2	614	11 Q7TMT6	Q7TMT6 mus musculu
38	254.5	37.3	613	11 Q8VXC7	Q8VXC7 mus musculu
39	252.5	37.0	168	11 Q8VDC9	Q8VDC9 mus musculu
40	252	36.9	112	4 Q9UCP3	Q9UCP3 mus musculu
41	250.5	36.7	147	11 Q9Z5S3	Q9Z5S3 mus musculu
42	250	36.6	159	4 Q96Q50	Q96Q50 mus musculu
43	248	36.3	479	11 Q95M22	Q95M22 mus musculu
44	247.5	36.2	143	11 Q92400	Q92400 mus musculu
45	247	36.2	119	4 Q9UL73	Q9UL73 mus musculu

## ALIGNMENTS

## RESULT 1

Q96BB9 ID Q96BB9 PRELIMINARY; PRT; 597 AA.

AC Q96BB9; 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Hypothetical protein.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=B-cell;  
RA Strausberg R.;  
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL, BC015760; AAH15760.1; -  
DR InterPro; IPR007110; IG\_1ike.  
DR InterPro; IPR003006; IG\_MHC.  
DR InterPro; IPR003596; IG\_V.  
DR Pfam; PF00047; IG\_5.  
DR SMART; SM00406; IG\_1.  
DR PROSITE; PSS0835; IG\_LIKE; 5.  
DR PROSITE; PS00290; IG\_MHC; 3.  
KW Hypothetical protein.  
SQ SEQUENCE 597 AA; 65039 MW; 4FCA3ADBECE263D9 CRC64;

Query Match 54.8%; Score 374; DB 4; Length 597;  
Best local Similarity 65.6%; Pred. No. 4.2e-29;  
Matches 82; Conservative 9; Mismatches 30; Indels 4; Gaps 3;

QY 1 QVQLQESGGGLVQAGGSLRLISCAASGISIHITMGVYRQTPGTERDVATIQ-DGGSTNY 59  
DB 20 EVQLTSGGGLVQPGGSLRLISCAASGFSSYAMWVRQAPGKLEWVAISGGSTNY 79  
QY 60 ADSVKRFTISRNINLTVYQNSLKPEDTAVYHGNADVPKRRS-RL--ELMGQCTL 116  
DB 80 ADSVKRFTISRNINLTVYQNSLKPEDTAVYHGNADVPKRRS-RL--ELMGQCTL 139

QY	117	VTVSS	121
Db	140	VTVSS	144

RESULT 2	
Q9UL72	
ID Q9UL72	PRELIMINARY; PRT; 118 AA

DT 01-MAY-2000 (TREMblrel, 13, Created)  
DT 01-MAY-2000 (TREMblrel, 13, last sequence update)  
DT 01-OCT-2003 (TREMblrel, 25, last annotation update)  
DE Myosin-reactive immunoglobulin heavy chain variable region  
DE (Fragment).  
OS Homo sapiens (human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
CX NCBI\_TaxID=9606;  
[1]  
RN  
RP SEQUENCE FROM N.A.  
RX MEDLINE=8277139; PubMed=9614934;  
RA Wang X., Liu B., Van der Werf P.L., Kalis N.N., Berney S.M.,  
RA Young D.C.;  
RT "Myosin-reactive autoantibodies in rheumatic carditis and normal  
RT fetus";  
RL Clin. Immunol. Immunopathol. 87:184-192(1998).  
DR EMBL; AF035042; ABDS6878.1; -.  
DR PIR; S21205; S21205.  
DR HSSP; P01772; 2FB4.  
DR InterPro; IPR007110; IG-like.  
DR InterPro; IPR03596; IG\_V.  
DR Pfam; PF00047; IG; 1.  
DR SMART; SM00405; IGV; 1.  
DR PROSITE; PSS0835; IG-LIKE; 1.  
DR NON\_TER 1  
FT NON\_TER 1  
SQ SEQUENCE 118 AA; 12672 MW; B4DIAS944B2DSCCA CXC64;

RESULT 3		
Q93XA4		
ID	Q93XA4	PRELIMINARY; PRT; 487 AA.
AC	Q93XA4;	
DT	01-JUN-2001 (TrEMBLrel. 17, Created)	
DT	01-JUN-2001 (TrEMBLrel. 17, last sequence update)	
DT	01-OCT-2003 (TrEMBLrel. 25, last annotation update)	
DE	Hypothetical protein.	
OS	Mus musculus (Mouse).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
OX	NCBI_TaxID=10090;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RA	Strausberg R.;	
RL	Submitted (MAR-2001) to the EMBL/GenBank/DDB databases.	
QR	EMBL; BC004786; AAH04786.1; -.	

DR HSSP; P01810; 2F3J.  
DR InterPro; IPR007110; IG-like.  
DR InterPro; IPR003006; IG\_MHC.  
DR InterPro; IPR003586; IG\_v.  
DR Pfam; PF00047; IG\_4.  
DR SMART; SM00406; IGV; 1.  
DR PROSITE; PS00835; IG\_LIKE; 4.  
DR PROSITE; PS00290; IG\_MHC; 2.  
KW Hypothetical protein  
SQ SEQUENCE 487 AA; 52554 MW; 7DC8E6DB333077B CRC64;

[illegible]



Db 140 TTVVSSASPTSPKVP 156

## RESULT 5

09UL90 PRELIMINARY; PRT; 113 AA.

AC 09UL90; 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
 DE Myosin-reactive immunoglobulin heavy chain variable region  
 (Fragment)  
 OS Homo sapiens (Human)  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 NCBI\_TaxID=9606;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98277139; PubMed=9614934;  
 RA Wu X., Liu B., Van der Merwe P.L., Kalis N.N., Berney S.M.,  
 Young D.C.;  
 RT "Myosin-reactive autoantibodies in rheumatic carditis and normal  
 fetus";  
 RL Clin. Immunol. Immunopathol. 87:184-192(1998).  
 DR EMBL: AF035024; AAD56260.1; -.  
 DR HSSP; P01772; 2FB4.  
 DR InterPro; IPR007110; Ig-like.  
 DR InterPro; IPR003596; Ig\_v.  
 DR Pfam; PF00047; Ig; 1.  
 DR SMART; SM00406; IGV; 1.  
 DR PROSITE; PS50835; IG-LIKE; 1.  
 DR SMART; SM00406; IGV; 1.  
 DR PROSITE; PS50835; IG-LIKE; 1.  
 FT NON TER 1  
 FT NON TER 113  
 SO SEQUENCE 113 AA; 12437 MW; ED57FDD19086D07F CRC64;

Query Match 52.4%; Score 358; DB 4; Length 113;  
 Best Local Similarity 63.9%; Pred. No. 2e-28;  
 Matches 78; Conservative 9; Mismatches 25; Indels 10; Gaps 2;

QY 1 QVQLQESGGGLVQAGSIRLSCAASGISHTMGWYRQTPGERDVATIQ-DGSGTNY 59  
 DB 1 EVQLVESGGGVQPGGSLRLSCAASGFTFSYGMHWVRQAPGKLEWAVFIRYDGSNKYY 60  
 QY 60 ADSVKGRTTISRDNILNTVYLQMSLKPEDTAVYHCNADVPRYTSRYLEIMGGTLTV 119  
 DB 61 ADSVKGRTTISRDNILNTVYLQMSLKPEDTAVYHCNADVPRYTSRYLEIMGGTLTV 111  
 QY 120 SS 121  
 DB 112 SS 113

## RESULT 6

08WUK1 PRELIMINARY; PRT; 613 AA.

AC 08WUK1; 01-MAR-2002 (TREMBLrel. 20, Created)  
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)  
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
 DE Hypothetical protein.  
 OS Homo sapiens (Human)  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 NCBI\_TaxID=9606;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC Tissue=Testis;  
 RA Strausberg R.;  
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC020240; AAH20240.1; -.  
 DR PIR; P10120; P10120.  
 DR PIR; S15590; S15590.

DR InterPro; IPR007110; Ig-like.  
 DR InterPro; IPR003006; Ig\_MHC.  
 DR InterPro; IPR003596; Ig\_v.  
 DR Pfam; PF00047; Ig; 5.  
 DR SMART; SM00406; IGV; 1.  
 DR PROSITE; PS50835; IG-LIKE; 5.  
 DR PROSITE; PS00290; IG\_MHC; 3.  
 KM Hypothetical protein.  
 SO SEQUENCE 613 AA; 67296 MW; 60C7F5950671E315 CRC64;

Query Match 51.0%; Score 348.5; DB 4; Length 613;  
 Best Local Similarity 61.1%; Pred. No. 1.6e-26;  
 Matches 77; Conservative 11; Mismatches 27; Indels 11; Gaps 3;

QY 1 QVQLQESGGGLVQAGSIRLSCAASGISHTMGWYRQTPGERDVATIQ-DGSGTNY 59  
 DB 20 QVQLVESGGGVQPGGSLRLSCAASGFTFSYGMHWVRQAPGKLEWAVISDGSNKYY 79  
 QY 60 ADSVKGRTTISRDNILNTVYLQMSLKPEDTAVYHCNADVPRYTSRYLEIMGGT 115  
 DB 80 ADSVKGRTTISRDNILNTVYLQMSLKPEDTAVYHCNADVPRYTSRYLEIMGGT 133  
 QY 116 LVTWSS 121  
 DB 134 MTWSS 139

## RESULT 7

09UL71 PRELIMINARY; PRT; 121 AA.

AC 09UL71; 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
 DE Myosin-reactive immunoglobulin heavy chain variable region  
 (Fragment).  
 OS Homo sapiens (Human)  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 NCBI\_TaxID=9606;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98277139; PubMed=9614934;  
 RA Wu X., Liu B., Van der Merwe P.L., Kalis N.N., Berney S.M.,  
 Young D.C.;  
 RT "Myosin-reactive autoantibodies in rheumatic carditis and normal  
 fetus";  
 RL Clin. Immunol. Immunopathol. 87:184-192(1998).  
 DR EMBL; AF035043; AAD56279.1; -.  
 DR HSSP; P01772; 2FB4.  
 DR InterPro; IPR007110; Ig-like.  
 DR InterPro; IPR003596; Ig\_v.  
 DR Pfam; PF00047; Ig; 1.  
 DR SMART; SM00406; IGV; 1.  
 DR PROSITE; PS50835; IG-LIKE; 1.  
 FT NON TER 1  
 FT NON TER 121  
 SO SEQUENCE 121 AA; 13154 MW; 2F045CCFA5D50736 CRC64;

Query Match 50.7%; Score 346; DB 4; Length 121;  
 Best Local Similarity 60.5%; Pred. No. 3.6e-27;  
 Matches 75; Conservative 15; Mismatches 28; Indels 6; Gaps 3;

QY 1 QVQLQESGGGLVQAGSIRLSCAASGISHTMGWYRQTPGERDVATIQ-DGSGTNY 59  
 DB 1 EVQLVESGGGVQPGGSLRLSCAASGFTFSYGMHWVRQAPGKLEWAVISDGSNKYY 60  
 QY 60 ADSVKGRTTISRDNILNTVYLQMSLKPEDTAVYHCNADVPRYTSRYLEIMGGTLTV 117  
 DB 61 ADSVKGRTTISRDNILNTVYLQMSLKPEDTAVYHCNADVPRYTSRYLEIMGGTLTV 117  
 QY 118 TWSS 121

Db 118 TVSS 121

## RESULT 8

Q91207 PRELIMINARY; PRT; 486 AA.

AC Q91207; 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Hypothetical protein.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Strausberg R.;  
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC010324; AAH10324.1; -  
 DR InterPro; IPR007110; IG\_1like.  
 DR InterPro; IPR003006; IG\_MHC.  
 DR InterPro; IPR003596; IG\_V.  
 DR Pfam; PF00047; IG\_1.  
 DR SMART; SM00406; IGV\_1.  
 DR PROSITE; PS00835; IG\_LIKE; 4.  
 DR PROSITE; PS00290; IG\_MHC; 2.  
 KW Hypothetical protein  
 SQ SEQUENCE 486 AA; 52682 MW; 4FF835125DA870B CRC64;

Query Match 50.1%; Score 346; DB 11; Length 486;  
 Best Local Similarity 55.9%; Pred. No. 2.2e-26;  
 Matches 71; Conservative 16; Mismatches 32; Indels 8; Gaps 2;

QY 1 QVQLDSGGGLVQAGSLRLSCAASGSIHTMGWVRQTPGERDVATIDGGSTNYA 60  
 DB 20 EVLVESGGGLVPGGSLRLSCVSGFSTISDMKWRQTPERRLRWVAATISGGNTYYP 79  
 QY 61 DSVKGRFTISRDNILNTVYLQWNSLKEPDTAYHCNADVRP-----YRTSRYLEMGQGT 115  
 DB 80 DNVKGRFTYSRDAKTYLQWNSLKSSEDTAYYC---VRPFIPIYVYSGSYFDSMGQGT 136  
 QY 116 LVTVSSSE 122  
 DB 137 TIVVSSSE 143

## RESULT 9

Q91XEL PRELIMINARY; PRT; 480 AA.

AC Q91XEL; 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Hypothetical protein (Fragment).  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Strausberg R.;  
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC010798; AAH10798.1; -  
 DR InterPro; IPR007110; IG\_1like.  
 DR InterPro; IPR003006; IG\_MHC.  
 DR InterPro; IPR003596; IG\_V.  
 DR Pfam; PF00047; IG\_1.  
 DR SMART; SM00406; IGV\_1.  
 DR PROSITE; PS00835; IG\_LIKE; 4.  
 DR PROSITE; PS00290; IG\_MHC; 2.  
 KW Hypothetical protein.

FT NON TER 1 1  
 SQ SEQUENCE 480 AA; 51936 MW; 20B9234EF2B41ED CRC64;

Query Match 50.1%; Score 342; DB 11; Length 480;  
 Best Local Similarity 61.0%; Pred. No. 5.4e-26;  
 Matches 75; Conservative 10; Mismatches 32; Indels 6; Gaps 2;

QY 2 VQLDSGGGLVQAGSLRLSCAASGSIHTMGWVRQTPGERDVATIDGG-STNYA 60  
 DB 20 VKLVESGGGLVPGGSLRLSCAASGTFPNSYMSWRQTPERRLRWVAATISGGVATHP 79  
 QY 61 DSVKGRFTISRDNILNTVYLQWNSLKEPDTAYHCNADVRPRTSRYLEMGQGTLYYS 120  
 DB 80 DSMKGRFTISRDAQNTVILQWNTSLNSSEDTAYVYCTRG----DWYFDVWGAGTIVVS 134  
 QY 121 SEP 123  
 DB 135 SEP 137

## RESULT 10

Q920E7 PRELIMINARY; PRT; 119 AA.

AC Q920E7; 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Pterin-mimicking anti-idiotope heavy chain variable region (Fragment).  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Atkin J.D., Iape A., Jennings I.G., Horatlis O., Cotton R.G.H.;  
 RT "Definition of the idiotope of Pterin-Mimicking Antibodies Expressed in Mammalian Cells";  
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF307937; AAU09421.1; -  
 DR PIR; C25913; C25913.  
 DR InterPro; IPR007110; IG\_1like.  
 DR InterPro; IPR003596; IG\_V.  
 DR Pfam; PF00047; IG\_1.  
 DR SMART; SM00406; IGV\_1.  
 DR PROSITE; PS00835; IG\_LIKE; 1.  
 DR NON TER 1  
 FT NON TER 119  
 SQ SEQUENCE 119 AA; 13025 MW; FE6904044381CA7C CRC64;

Query Match 49.8%; Score 340; DB 11; Length 119;  
 Best Local Similarity 60.8%; Pred. No. 1.4e-26;  
 Matches 76; Conservative 10; Mismatches 29; Indels 10; Gaps 3;

QY 1 QVQLDSGGGLVQAGSLRLSCAASGSIHTMGWVRQTPGERDVATIDGG-STNY 59  
 DB 1 EVLVESGGDLVPGGSLRLSCAASGTFPNSYMSWRQTPERRLRWVAATISGGGYTYA 60  
 QY 60 ADSVKGFTISRDNILNTVYLQWNSLKEPDTAYHCNADVRPRTSRYLEMGQGT 116  
 DB 61 PDVSKGRFTISRDAKTYLQWNSLKSSEDTAYYC-----ARHSDYVGFAYMGQGT 114

QY 117 TVVSS 121  
 DB 115 TVVSA 119

## RESULT 11

Q9UL88 PRELIMINARY; PRT; 131 AA.

ID Q9UL88; 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)

DT 01-OCT-2003 (TREMblrel. 25, last annotation update)  
 DE Myosin-reactive Immunoglobulin heavy chain variable region  
 DE (Fragment).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 CX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96277139; PubMed=9614934;  
 RA Wu X., Liu B., Van der Merwe P.L., Kalis N.N., Berney S.M.,  
 RA Young D.C.;  
 RT "Myosin-reactive autoantibodies in rheumatic carditis and normal  
 RT fetus";  
 RL Clin. Immunol. Immunopathol. 87:184-192(1998).  
 DR EMBL; AF035026; AAD56262.1; -.  
 DR PIR; S21205; S21205.  
 DR HSSP; P01810; 2FBJ.  
 DR InterPro; IPR007110; Ig-like.  
 DR InterPro; IPR003596; Ig\_v.  
 DR Pfam; PF00047; Ig\_1.  
 DR SMART; SM00406; IgV\_1.  
 DR PROSITE; PS00835; IG\_LIKE; 1.  
 FT NON\_TER  
 FT NON\_TER  
 SQ SEQUENCE 131 AA; 14142 MW; 9657D66B375DEA0 CRC64;

Query Match 49.6%; Score 339; DB 4; Length 131;  
 Best Local Similarity 57.9%; Pred. No. 2e-26;  
 Matches 77; Conservative 11; Mismatches 31; Indels 14; Gaps 3;

QY 1 QVQLQESGGGLVQAGSRLRSCAASGSGISHTMGWRQTPGERDVAATIQ--DGGST 57  
 DB 1 EVQLVESGGGLVPGKSLRLSCAASGFFPKAMSWVRQAGKGLVWGRISKNTDGT 60  
 CY 58 NYADSVKRFITSRDNIINTVYLQNSLKPEDTAVYHCNADV-----RPRTRSYL 108  
 DB 61 DYAAEVKRLTISRDSKNTLYLRNKSITKEDTAVYCTGTITMIVITTSKRTS--F 118  
 QY 109 ELMGGTLVTSS 121  
 DB 119 EYWGQTLVTSS 131

## RESULT 12

ID Q9Y509 PRELIMINARY; PRT; 147 AA.  
 AC Q9Y509;

DT 01-NOV-1999 (TREMblrel. 12, Created)  
 DT 01-NOV-1999 (TREMblrel. 12, Last sequence update)  
 DT 01-OCT-2003 (TREMblrel. 25, Last annotation update)  
 DE Vh3 protein (Fragment).  
 GN Vh3.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 CX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96071149; PubMed=7475288;  
 RA Cao J., Vesicic R.A., Rettig M.B., Hong C.H., Kim A., Lee J.C.,  
 RA Lieberstein A.K., Berenson J.R.;  
 RT "A CD10-positive subset of malignant cells is identified in multiple  
 RT myeloma using PCR with patient-specific immunoglobulin gene primers";  
 RL Leukemia 9:1948-1953(1995).  
 DR EMBL; S80860; AAD14339.1; -.  
 DR HSSP; P01772; 2FBJ.  
 DR GO; GO:0005887; C:intracellular to plasma membrane; NAS.  
 DR GO; GO:0036489; F:immunoglobulin receptor activity; NAS.  
 DR InterPro; IPR007110; Ig-like.  
 DR InterPro; IPR003596; Ig\_v.  
 DR Pfam; PF00047; Ig\_1.

DR SMART; SM00406; IgV\_1.  
 DR PROSITE; PS00835; IG\_LIKE; 1.  
 FT NON\_TER  
 SQ SEQUENCE 147 AA; 15768 MW; 8489FCAA7BC925C CRC64;

Query Match 49.6%; Score 339; DB 4; Length 147;  
 Best Local Similarity 58.1%; Pred. No. 2.4e-26;  
 Matches 79; Conservative 9; Mismatches 42; Indels 6; Gaps 3;

QY 1 QVQLQESGGGLVQAGSRLRSCAASGSGISHTMGWRQTPGERDVAATIQDGGSTN-Y 59  
 DB 1 QVQLVESGGGLVPGKSLRLSCAASGFTSTYGMSSVRQAGKGLVWVALISYDSTQY 60  
 QY 60 ADSVKGKRFITSRDNIINTVYLQNSLKPEDTAVYHCNADVPRTRSYR---LEMGQGT 115  
 DB 61 AGSVKGKRFITSRDNIINTVYLQNSLKPEDTAVYHCNADVPRTRSYR---LEMGQGT 120  
 QY 116 LVTYSS-EPKTPKPPQ 130  
 DB 121 LVTYSSASTKGPSPVP 136

## RESULT 13

ID Q91WP5 PRELIMINARY; PRT; 479 AA.  
 AC Q91WP5;

DT 01-DEC-2001 (TREMblrel. 19, Created)  
 DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)  
 DT 01-OCT-2003 (TREMblrel. 25, Last annotation update)  
 DE Hypothetical protein.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 CX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA TISUB=Colon;  
 RA Strausberg R.;  
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC013656; AAH13656.1; -.  
 DR InterPro; IPR007110; Ig-like.  
 DR InterPro; IPR003006; Ig\_MHC.  
 DR InterPro; IPR003596; Ig\_v.  
 DR Pfam; PF00047; Ig\_4.  
 DR SMART; SM00406; IgV\_1.  
 DR PROSITE; PS00835; IG\_LIKE; 4.  
 DR PROSITE; PS00290; IG\_MHC; 2.  
 KW Hypothetical protein.  
 SQ SEQUENCE 479 AA; 51603 MW; ECB2D0877748584F CRC64;

Query Match 49.3%; Score 337; DB 11; Length 479;  
 Best Local Similarity 57.3%; Pred. No. 1.7e-25;  
 Matches 71; Conservative 19; Mismatches 26; Indels 8; Gaps 3;

QY 1 QVQLQESGGGLVQAGSRLRSCAASGSGISHTMGWRQTPGERDVAATIQDGGSTNY 59  
 DB 20 EVQLVESGGGLVPGKSLRLSCAASGFTSTYGMSSVRQAGKGLVWVALISYDSTQY 79  
 QY 60 ADSVKGKRFITSRDNIINTVYLQNSLKPEDTAVYHCNADVPRTRSYRLEMGQGLTV 119  
 DB 80 SDPMKRFITSRDNIINTVYLQNSLKPEDTAVYHCNADVPRTRSYRLEMGQGLTV 132  
 QY 120 SSSP 123  
 DB 133 SSSP 136

## RESULT 14

ID Q8WJ38 PRELIMINARY; PRT; 573 AA.  
 AC Q8WJ38;

DT 01-MAR-2002 (TREMblrel. 20, Created)  
 DT 01-MAR-2002 (TREMblrel. 20, Last sequence update)



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd

OM protein - protein search, using sw model

```
Run on:      October 5, 2004, 07:04:10 ; Search time 98.903 Seconds
              (without alignments)
              371.381 Million cell updates/sec
```

Title: US-09-805-290A-21  
Perfect score: 683  
Sequence: 1 QVQLQESGGGLVQAGSLRL...WGQGLTVVSSEPKTKPQP 130

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

```
Post-processing:  Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries
```

```
Database :
1: Geneset_qp29a04:*
A: Geneset_qp1980s:*
1: Geneset_qp1980s:*
2: Geneset_qp1990s:*
3: Geneset_qp2000s:*
4: Geneset_qp2001s:*
5: Geneset_qp2002s:*
6: Geneset_qp2003s:*
7: Geneset_qp2004s:*
8: Geneset_qp2005s:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	653	100.0	130	4	AAE10554	AAe10554 HPL inhib
2	689	96.5	130	4	AAE10552	AAe10552 HPL inhib
3	617	90.3	130	4	AAE10559	AAe10559 HPL inhib
4	520.5	76.2	131	4	AAE10555	AAe10555 HPL inhib
5	513	75.1	130	4	AAE10556	AAe10556 HPL inhib
6	500	73.2	130	4	AAE10557	AAe10557 HPL inhib
7	492.5	72.1	129	4	AAE10553	AAe10553 HPL inhib
8	489	71.6	152	4	AAE05283	AAe05283 Anti -pote
9	488	71.6	124	4	AAE10563	AAe10563 HGL inhib
10	480.5	70.4	131	5	ABG30620	ABg30620 Immuno
11	476	69.7	124	4	AAE10566	AAe10566 HPL inhib
12	467.5	68.4	129	4	AAE10558	AAe10558 HPL inhib
13	466.5	68.3	149	4	AAE05288	AAe05288 Anti -pote
14	464.5	68.0	129	4	AAE10561	AAe10561 HGL inhib
15	463	67.8	130	4	AAE10567	AAe10567 HGL inhib
16	460.5	67.4	129	4	AAE10564	AAe10564 HGL inhib
17	458	67.1	128	4	AAE10565	AAe10565 HGL inhib
18	447	65.4	190	7	ABR62883	ABr62883 Liama ant
19	446.5	65.4	129	4	AAE10551	AAe10551 HPL inhib
20	445	65.2	190	7	ABR62882	ABr62882 Liama ant
21	444.5	65.1	377	4	AAE05286	AAe05286 Liama HCV
22	444.5	65.1	383	4	AAE05287	AAe05287 Liama HCV
23	443	64.9	119	4	AAE67778	AAe67778 Amino aci
24	443	64.9	119	4	AAE67777	AAe67777 Amino aci
25	443	64.9	190	7	ABR62881	ABr62881 Liama ant

## ALIGNMENTS

27	441.5	64.6	118	2	AAV39819	Llama ant	AAV39819	Llama ant
26	441.5	64.6	204	2	AAV41167	Llama ant	AAV41167	Llama ant
28	441	64.6	190	7	ABR62879	Llama ant	ABR62879	Llama ant
29	440	64.4	190	7	ABR62880	Llama ant	ABR62880	Llama ant
30	440	64.4	198	7	ABR62878	Llama ant	ABR62878	Llama ant
31	436.5	63.9	211	2	AAV41172	Llama Vhh	AAV41172	Llama Vhh
32	435	63.5	124	2	ABG30618	Immunoglo	ABG30618	Immunoglo
33	433	63.4	119	4	ABR67779	Anti-pocta	ABR67779	Anti-pocta
34	432.5	63.3	153	4	AAE05282	Llama Vhh	AAE05282	Llama Vhh
35	426.5	62.2	225	2	AAV41165	Llama Vhh	AAV41165	Llama Vhh
36	424.5	62.4	205	2	AAV41173	Llama Vhh	AAV41173	Llama Vhh
37	423	61.9	124	4	AAE10562	Hgt. inhib	AAE10562	Hgt. inhib
38	423	61.9	194	7	ABR62885	Llama ant	ABR62885	Llama ant
39	422.5	61.9	127	5	ABG30617	Mouse circ	ABG30617	Mouse circ
40	420.5	61.6	133	5	AAV39812	Immunoglo	AAV39812	Immunoglo
41	419.5	61.4	124	2	AAV29615	Llama ant	AAV29615	Llama ant
42	416	60.9	125	2	AAV39817	Llama ant	AAV39817	Llama ant
43	414.5	60.7	193	7	ABR62884	Llama ant	ABR62884	Llama ant
44	414	60.6	130	4	ABE10560	Hgt. inhib	ABE10560	Hgt. inhib
45	410.5	60.1	183	2	AAV41166	Llama Vhh	AAV41166	Llama Vhh

XX	AAE10554	standard; peptide; 130 AA.
XX	AAE10554	
XX	10-DEC-2001	(first entry)
DE	HPL inhibiting VHH fragment, HPL #14 from llama species.	
XX	Llama antibody; camelid; antineoplastic; heavy chain variable domain; VHH;	
KM	human dietary enzyme inhibitor; medicament; human pancreatic lipase; HPL	
KM	food; human gastric lipase; HGL; cosmetic control; body weight.	
XX	Llama sp.	
OS		
XX	Key	Location/Qualifiers
FH	Region	31..35
FT		/label=CDR1
FT		/note="Complementarity determining region 1"
FT	Region	50..54
FT		/label=CDR2
FT	Region	/note="Complementarity determining region 2"
FT		98..110
FT		/label=CDR3
XX		/note="Complementarity determining region 3"
XX	EP134231-A1.	
XX	19-SEP-2001.	
XX	20-FEB-2001; 2001EP-00200703.	
XX	14-MAR-2000; 2000EP-00200930.	
XX	(UNIL ) UNILEVER NV.	
XX	(UNIL ) UNILEVER PLC.	
XX	Bezemer S, Van De Burg M, De Haard JTW, Tareilus E,	
XX	WPI: 2001-572718/65.	
XX	New antibody or its fragments for inhibiting human dietary enzymes,	
XX	useful for cosmetic control of body weight of human beings, comprises	
XX	heavy chain variable domain derived from immunoglobulin naturally devoid	
XX	of light chains.	
XX	Example 2; Page 9; 37pp; English.	

XX The patent discloses antibodies or their fragments comprising a heavy  
CC chain variable domain (VH) derived from an immunoglobulin naturally  
CC devoid of light chains specific for inhibiting human dietary enzymes. The  
CC antibodies of the invention are useful for the preparation of medicaments  
CC or food for inhibiting the activity of one or more human dietary enzymes  
CC especially human pancreatic lipase (HPL) or human gastric lipase (HGL)  
CC which are useful for the cosmetic control of body weight of human beings.  
CC The present peptide sequence is HPL inhibiting VH fragment, HPL #14 from  
CC llama (camelid) species  
XX  
SQ Sequence 130 AA;  
Query Match 100.0%; Score 683; DB 4; Length 130;  
Best Local Similarity 100.0%; Pred. No. 4.8e-54;  
Matches 130; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 QVQLQESGGGLVQAGSRLRSCAASGISGSIHTMGWYRQTPGERDVAITIDGGSTNYA 60  
DB 1 QVQLQESGGGLVQAGSRLRSCAASGISGSIHTMGWYRQTPGERDVAITIDGGSTNYA 60  
QY 61 DSVKGRFTISRDNILNTVYLQNMNLSKPEDTAVYHCNADVRPRTSRYLELMOGTLYVTS 120  
DB 61 DSVKGRFTISRDNILNTVYLQNMNLSKPEDTAVYHCNADVRPRTSRYLELMOGTLYVTS 120  
QY 121 SEPTPKPKOP 130  
DB 121 SEPTPKPKOP 130  
RESULT 2  
AAE10552  
ID AAE10552 standard; peptide; 130 AA.  
AC AAE10552;  
XX  
DT 10-DEC-2001 (first entry)  
XX  
DE HPL inhibiting VH fragment, HPL #12 from llama species.  
XX  
KW llama antibody; camelid; anorectic; heavy chain variable domain; VH;  
KW human dietary enzyme inhibitor; medicament; human pancreatic lipase; HPL;  
KW food; human gastric lipase; HGL; cosmetic control; body weight.  
XX  
OS Lama sp.  
XX  
FH Key Location/Qualifiers  
FT Region 31..35  
FT /label= CDR1  
FT /note= "Complementarity determining region 1"  
FT 50..64  
FT /label= CDR2  
FT /note= "Complementarity determining region 2"  
FT 98..110  
FT /label= CDR3  
FT /note= "Complementarity determining region 3"  
XX  
PN EPI134231-A1.  
XX  
PD 19-SEP-2001.  
XX  
PF 20-FEB-2001; 2001EP-00200703.  
XX  
PR 14-MAR-2000; 2000EP-00200930.  
XX  
PA (UNITL ) UNILEVER NV.  
PA (UNITL ) UNILEVER PLC.  
XX  
PI Bezemer S, Van De Burg M, De Haard JWM, Tareilus E;  
XX  
DR WPI; 2001-572718/65.  
XX  
PT New antibody or its fragments for inhibiting human dietary enzymes,

PT useful for cosmetic control of body weight of human beings, comprises  
PT heavy chain variable domain derived from immunoglobulin naturally devoid  
PT of light chains.  
XX  
PS Example 2; Page 9; 37pp; English.  
XX  
XX The patent discloses antibodies or their fragments comprising a heavy  
CC chain variable domain (VH) derived from an immunoglobulin naturally  
CC devoid of light chains specific for inhibiting human dietary enzymes. The  
CC antibodies of the invention are useful for the preparation of medicaments  
CC or food for inhibiting the activity of one or more human dietary enzymes  
CC especially human pancreatic lipase (HPL) or human gastric lipase (HGL)  
CC which are useful for the cosmetic control of body weight of human beings.  
CC The present peptide sequence is HPL inhibiting VH fragment, HPL #12 from  
CC llama (camelid) species  
XX  
SQ Sequence 130 AA;  
Query Match 96.5%; Score 659; DB 4; Length 130;  
Best Local Similarity 96.2%; Pred. No. 7.1e-52;  
Matches 125; Conservative 2; Mismatches 3; Indels 0; Gaps 0;  
QY 1 QVQLQESGGGLVQAGSRLRSCAASGISGSIHTMGWYRQTPGERDVAITIDGGSTNYA 60  
DB 1 QVQLQESGGGLVQAGSRLRSCAASGISGSIHTMGWYRQTPGERDVAITIDGGSTNYA 60  
QY 61 DSVKGRFTISRDNILNTVYLQNMNLSKPEDTAVYHCNADVRPRTSRYLELMOGTLYVTS 120  
DB 61 DSVKGRFTISRDNILNTVYLQNMNLSKPEDTAVYHCNADVRPRTSRYLELMOGTLYVTS 120  
QY 121 SEPTPKPKOP 130  
DB 121 SEPTPKPKOP 130  
RESULT 3  
AAE10559  
ID AAE10559 standard; peptide; 130 AA.  
AC AAE10559;  
XX  
DT 10-DEC-2001 (first entry)  
XX  
DE HPL inhibiting VH fragment, HPL #30 from llama species.  
XX  
KW llama antibody; camelid; anorectic; heavy chain variable domain; VH;  
KW human dietary enzyme inhibitor; medicament; human pancreatic lipase; HPL;  
KW food; human gastric lipase; HGL; cosmetic control; body weight.  
XX  
OS Lama sp.  
XX  
FH Key Location/Qualifiers  
FT Region 31..35  
FT /label= CDR1  
FT /note= "Complementarity determining region 1"  
FT 49..64  
FT /label= CDR2  
FT /note= "Complementarity determining region 2"  
FT 98..110  
FT /label= CDR3  
FT /note= "Complementarity determining region 3"  
XX  
PN EPI134231-A1.  
XX  
PD 19-SEP-2001.  
XX  
PF 20-FEB-2001; 2001EP-00200703.  
XX  
PR 14-MAR-2000; 2000EP-00200930.  
XX  
PA (UNITL ) UNILEVER NV.  
PA (UNITL ) UNILEVER PLC.  
XX  
XX

PI Bezemer S, Van De Burg M, De Haard JTW, Tareilus E;  
 XX WPI; 2001-572718/65.  
 XX  
 XX New antibody or its fragments for inhibiting human dietary enzymes,  
 PT useful for cosmetic control of body weight of human beings, comprises  
 PT heavy chain variable domain derived from immunoglobulin naturally devoid  
 PT of light chains.  
 XX  
 XX Example 2; Page 10; 37pp; English.  
 XX  
 XX The patent discloses antibodies or their fragments comprising a heavy  
 CC chain variable domain (VH) derived from an immunoglobulin naturally  
 CC devoid of light chains specific for inhibiting human dietary enzymes. The  
 CC antibodies of the invention are useful for the preparation of medicaments  
 CC or food for inhibiting the activity of one or more human dietary enzymes  
 CC especially human pancreatic lipase (HPL) or human gastric lipase (HGL)  
 CC which are useful for the cosmetic control of body weight of human beings.  
 CC The present peptide sequence is HPL inhibiting VH fragment, HPL #30 from  
 CC llama (camelid) species  
 XX  
 SQ Sequence 130 AA;  
 Query Match 90.3%; Score 617; DB 4; Length 130;  
 Best Local Similarity 90.0%; Pred. No. 4,4e-48;  
 Matches 117; Conservative 4; Mismatches 9; Indels 0; Gaps 0;  
 QY 1 QVQLQESGGGLVQAGGSLRLSCAASGSIHTMGWYRQPTGTRDVVATIDGGSTNYA 60  
 DB 1 QVQLDSSGGGLVQAGGSLRLSCAASGSIQDHTMGWIRQTKGQPDVYATIQSGGSNTYA 60  
 QY 61 DSVKGRFTISRDNLTNTVYIQMNSLKPEDTAVYHCNADVPRTSRYLELWGCGTLVTVS 120  
 DB 61 DSVKGRFTISRDNLTNTVYIQMNDLKPEDTGVYWNADVPRYTSRYLELWGCGTLVTVF 120  
 QY 121 SEPKTKPKQP 130  
 DB 121 LEPKTKPKQP 130  
 RESULT 4  
 AAE10555  
 ID AAE10555 standard; peptide; 131 AA.  
 XX  
 XX AAE10555;  
 AC  
 XX  
 DT 10-DEC-2001 (first entry)  
 XX  
 DE HPL inhibiting VH fragment, HPL #15 from llama species.  
 XX  
 XX llama antibody; camelid; anorectic; heavy chain variable domain; VH;  
 KW human dietary enzyme inhibitor; medicament; human pancreatic lipase; HPL;  
 KW food; human gastric lipase; HGL; cosmetic control; body weight.  
 XX  
 OS Lama sp.  
 XX  
 FH Key Location/Qualifiers  
 FT 31..35  
 FT /label=CDR1  
 FT /note="Complementarity determining region 1"  
 FT 50..64  
 FT /label=CDR2  
 FT /note="Complementarity determining region 2"  
 FT 98..111  
 FT /label=CDR3  
 FT /note="Complementarity determining region 3"  
 FT Region  
 XX  
 XX EPI134231-A1.  
 XX  
 XX 19-SEP-2001.  
 XX  
 XX 20-FEB-2001; 2001EP-00200703.  
 XX

PR 14-MAR-2000; 2000EP-00200930.  
 XX  
 XX (UNITL) UNILEVER NV.  
 PA (UNITL) UNILEVER PLC.  
 XX  
 XX Bezemer S, Van De Burg M, De Haard JTW, Tareilus E;  
 PI WPI; 2001-572718/65.  
 DR  
 XX  
 XX New antibody or its fragments for inhibiting human dietary enzymes,  
 PT useful for cosmetic control of body weight of human beings, comprises  
 PT heavy chain variable domain derived from immunoglobulin naturally devoid  
 PT of light chains.  
 XX  
 XX Example 2; Page 9; 37pp; English.  
 XX  
 XX The patent discloses antibodies or their fragments comprising a heavy  
 CC chain variable domain (VH) derived from an immunoglobulin naturally  
 CC devoid of light chains specific for inhibiting human dietary enzymes. The  
 CC antibodies of the invention are useful for the preparation of medicaments  
 CC or food for inhibiting the activity of one or more human dietary enzymes  
 CC especially human pancreatic lipase (HPL) or human gastric lipase (HGL)  
 CC which are useful for the cosmetic control of body weight of human beings.  
 CC The present peptide sequence is HPL inhibiting VH fragment, HPL #15 from  
 CC llama (camelid) species  
 XX  
 SQ Sequence 131 AA;  
 Query Match 76.2%; Score 520.5; DB 4; Length 131;  
 Best Local Similarity 76.3%; Pred. No. 2,4e-39;  
 Matches 100; Conservative 12; Mismatches 18; Indels 1; Gaps 1;  
 QY 1 QVQLQESGGGLVQAGGSLRLSCAASGSIHTMGWYRQPTGTRDVVATIDGGSTNYA 60  
 DB 1 QVQLQESGGGLVQAGGSLRLSCAASGSIINVMGWFRAQPKQRLVASITSGGSNTYA 60  
 QY 61 DSVKGRFTISRDNLTNTVYIQMNSLKPEDTAVYHCNADVPRTSRYLELWGCGTLVTV 119  
 DB 61 DSVKGRFTISRDNLTNTVYIQMNSLKPEDTAVYHCNADVPRTSRYLELWGCGTLVTV 120  
 QY 120 SEPKTKPKQP 130  
 DB 121 SEPKTKPKQP 131  
 RESULT 5  
 AAE10556  
 ID AAE10556 standard; peptide; 130 AA.  
 XX  
 XX AAE10556;  
 AC  
 XX  
 DT 10-DEC-2001 (first entry)  
 XX  
 DE HPL inhibiting VH fragment, HPL #18 from llama species.  
 XX  
 XX llama antibody; camelid; anorectic; heavy chain variable domain; VH;  
 KW human dietary enzyme inhibitor; medicament; human pancreatic lipase; HPL;  
 KW food; human gastric lipase; HGL; cosmetic control; body weight.  
 XX  
 OS Lama sp.  
 XX  
 FH Key Location/Qualifiers  
 FT 31..35  
 FT /label=CDR1  
 FT /note="Complementarity determining region 1"  
 FT 50..64  
 FT /label=CDR2  
 FT /note="Complementarity determining region 2"  
 FT 98..110  
 FT /label=CDR3  
 FT /note="Complementarity determining region 3"  
 FT Region  
 XX  
 XX EPI134231-A1.  
 XX

```

XX 19-SEP-2001.
PD 20-FEB-2001, 2001EP-00200703.
XX 14-MAR-2000, 2000EP-00200930.
XX (UNIL ) UNILEVER NV.
XX (UNIL ) UNILEVER PLC.
XX Bezemer S, Van De Burg M, De Haard JW, Tareilus E;
XX WPI, 2001-572718/65.
XX
XX New antibody or its fragments for inhibiting human dietary enzymes,
XX useful for cosmetic control of body weight of human beings, comprises
XX heavy chain variable domain derived from immunoglobulin naturally devoid
XX of light chains.
XX
XX Example 2, Page 10; 37pp; English.
XX
XX The patent discloses antibodies or their fragments comprising a heavy
XX chain variable domain (VH) derived from an immunoglobulin naturally
XX devoid of light chains specific for inhibiting human dietary enzymes. The
XX antibodies of the invention are useful for the preparation of medicaments
XX or food for inhibiting the activity of one or more human dietary enzymes
XX especially human pancreatic lipase (HPL) or human gastric lipase (HGL)
XX which are useful for the cosmetic control of body weight of human beings.
XX The present peptide sequence is HPL inhibiting VH fragment, HPL #18 from
XX llama (camelid) species
XX
XX Sequence 130 AA;
SQ
Query Match 75.1%; Score 513; DB 4; Length 130;
Best Local Similarity 76.9%; Pred. No. 1.1e-38;
Matches 100; Conservative 10; Mismatches 20; Indels 0; Gaps 0;

QY 1 QVQLQESGGGLVQAGSLRLSCAASGSIHTMGVYRQTPGTERDVAITIDGGSTNYA 60
   |||||
Db 1 QVQLQESGGGLVQAGSLRLSCAASGSIHTMGVYRQTPGTERDVAITIDGGSTNYA 60
   |||||
QY 61 DSVKGRFTISRDNILNTVYLQNNSLKPEDTAVYHCNADVPYRTSRYLEMGQGLTVTS 120
   |||||
Db 61 DPKVGRFTISRDNKLTLYLQNNSLKPEDTAVYHCNADVPYRTSRYLEMGQGLTVTS 120
   |||||
QY 121 SEPTKPKQP 130
   |||||
Db 121 SEPTKPKQP 130
   |||||

RESULT 6
AAE10557
ID AAE10557 standard; peptide; 130 AA.
XX
XX AAE10557;
XX
XX 10-DEC-2001 (first entry)
XX
XX HPL inhibiting VH fragment, HPL #19 from llama species.
XX
XX Llama antibody; camelid; anorectic; heavy chain variable domain; VH;
XX human dietary enzyme inhibitor; medicament; human pancreatic lipase; HPL;
XX food; human gastric lipase; HGL; cosmetic control; body weight.
XX
XX Llama sp.
XX
XX Key Location/Qualifiers
XX Region 31..35 CDR1
XX /label= CDR1
XX /note= "Complementarity determining region 1"
XX 50..64
XX /label= CDR2
XX /note= "Complementarity determining region 2"

```

```

FT Region 98..110
FT /label= CDR3
FT /note= "Complementarity determining region 3"
XX
XX EPI134231-A1.
XX
XX 19-SEP-2001.
XX
XX 20-FEB-2001, 2001EP-00200703.
XX
XX 14-MAR-2000, 2000EP-00200930.
XX
XX (UNIL ) UNILEVER NV.
XX (UNIL ) UNILEVER PLC.
XX Bezemer S, Van De Burg M, De Haard JW, Tareilus E;
XX WPI, 2001-572718/65.
XX
XX New antibody or its fragments for inhibiting human dietary enzymes,
XX useful for cosmetic control of body weight of human beings, comprises
XX heavy chain variable domain derived from immunoglobulin naturally devoid
XX of light chains.
XX
XX Example 2, Page 10; 37pp; English.
XX
XX The patent discloses antibodies or their fragments comprising a heavy
XX chain variable domain (VH) derived from an immunoglobulin naturally
XX devoid of light chain specific for inhibiting human dietary enzymes. The
XX antibodies of the invention are useful for the preparation of medicaments
XX or food for inhibiting the activity of one or more human dietary enzymes
XX especially human pancreatic lipase (HPL) or human gastric lipase (HGL)
XX which are useful for the cosmetic control of body weight of human beings.
XX The present peptide sequence is HPL inhibiting VH fragment, HPL #19 from
XX llama (camelid) species
XX
XX Sequence 130 AA;
SQ
Query Match 73.2%; Score 500; DB 4; Length 130;
Best Local Similarity 73.8%; Pred. No. 1.7e-37;
Matches 96; Conservative 10; Mismatches 24; Indels 0; Gaps 0;

QY 1 QVQLQESGGGLVQAGSLRLSCAASGSIHTMGVYRQTPGTERDVAITIDGGSTNYA 60
   |||||
Db 1 QVQLQESGGGLVQAGSLRLSCAASGSIHTMGVYRQTPGTERDVAITIDGGSTNYA 60
   |||||
QY 61 DSVKGRFTISRDNILNTVYLQNNSLKPEDTAVYHCNADVPYRTSRYLEMGQGLTVTS 120
   |||||
Db 61 DSVKGRFTISRDNKLTLYLQNNSLKPEDTAVYHCNADVPYRTSRYLEMGQGLTVTS 120
   |||||
QY 121 SEPTKPKQP 130
   |||||
Db 121 SEPTKPKQP 130
   |||||

RESULT 7
AAE10553
ID AAE10553 standard; peptide; 129 AA.
XX
XX AAE10553;
XX
XX 10-DEC-2001 (first entry)
XX
XX HPL inhibiting VH fragment, HPL #13 from llama species.
XX
XX Llama antibody; camelid; anorectic; heavy chain variable domain; VH;
XX human dietary enzyme inhibitor; medicament; human pancreatic lipase; HPL;
XX food; human gastric lipase; HGL; cosmetic control; body weight.
XX
XX Llama sp.
XX
XX Key Location/Qualifiers
XX Region 31..35

```



```

FT /label= CDR1
FT /note= "Complementarity determining region 1"
FT Region
FT 50..64
FT /label= CDR2
FT /note= "Complementarity determining region 2"
FT Region
FT 98..109
FT /label= CDR3
FT /note= "Complementarity determining region 3"
XX
XX EPI134231-A1.
XX
XX 19-SEP-2001.
XX
XX 20-FEB-2001; 2001EP-00200703.
XX
XX 14-MAR-2000; 2000EP-00200930.
XX
XX (UNIL ) UNILEVER NV.
XX (UNIL ) UNILEVER PLC.
XX
XX Bezemer S, Van De Burg M, De Haard JMW, Tareilus B;
XX WPI; 2001-572716/65.
XX
XX New antibody or its fragments for inhibiting human dietary enzymes,
XX useful for cosmetic control of body weight of human beings, comprises
XX heavy chain variable domain derived from immunoglobulin naturally devoid
XX of light chains.
XX
XX Example 2; Page 9; 37pp; English.
XX
XX The patent discloses antibodies or their fragments comprising a heavy
XX chain variable domain (VH) derived from an immunoglobulin naturally
XX devoid of light chains specific for inhibiting human dietary enzymes. The
XX antibodies of the invention are useful for the preparation of medicaments
XX or food for inhibiting the activity of one or more human dietary enzymes
XX especially human pancreatic lipase (HPL) or human gastric lipase (HGL)
XX which are useful for the cosmetic control of body weight of human beings.
XX The present peptide sequence is HPL inhibiting VH fragment, HPL #13 from
XX llama (camelid) species
XX
XX Sequence 129 AA;
XX
XX Query Match 72.1%; Score 492.5; DB 4; Length 129;
XX Best Local Similarity 76.2%; Pred. No. 7.9e-37;
XX Matches 99; Conservative 9; Mismatches 21; Indels 1; Gaps 1;
XX
XX 1 QVQLQESGGGLVQAGGSLRLSCAASGSIHTMGWYRQTPGTERDVAATIDGGSTNYA 60
XX 1 QVQLQESGGGLVQAGGSLRLSCAASGTLISITIMDWYGTGKGRVLRGRTTAAAGSTNYA 60
XX
XX 61 DSVKGRFTISRDNIINTVYLQWNSLKPEDTAIVHCHADVPKRSRYIELMGQGLTVTS 120
XX 61 DSAKRFTISRDNAKNTVYLQWNSLKPEDTAIVYCNALITWDXS-VMDYWGQGTQVTVS 119
XX
XX 121 SEPKTKPKPOP 130
XX 120 SEPKTKPKPOP 129
XX
XX
XX RESULT 8
XX AAE05283
XX ID AAE05283 standard; protein; 153 AA.
XX
XX AAE05283;
XX
XX 18-SEP-2001 (first entry)
XX
XX Anti-potato SBEII (Clone46) VH region attached with myc and his6 tag.
XX
XX Potato; heavy chain immunoglobulin; pathogen resistance;
XX metabolism modulator; passive immunisation; heavy chain variable domain;
XX VH; anti-potato SBEII; starch branching enzyme; SBE A.

```

```

XX Solanum tuberosum.
XX OS Undifferentiated.
XX OS Chimeric.
XX PN EPI118669-A2.
XX
XX 25-JUL-2001.
XX
XX 08-DEC-2000; 2000EP-00310997.
XX
XX 17-DEC-1999; 99EP-00310188.
XX
XX (UNIL ) UNILEVER PLC.
XX (UNIL ) UNILEVER NV.
XX
XX Frenken LGJ, Van Der Logt CPE, Jobling SA, Teh Y;
XX WPI; 2001-427157/46.
XX N-PSDB; AAD10054.
XX
XX Modifying a plant to produce an antibody useful for increasing pathogen
XX resistance or to modulate metabolism comprises introducing a DNA sequence
XX encoding a heavy chain immunoglobulin linked to a peptide that targets a
XX cellular compartment.
XX
XX Example 1; Fig 12; 81pp; English.
XX
XX The present invention relates to a method for modifying a plant to
XX produce an antibody or an active fragment or derivative, or a protein
XX functional equivalent, in a cellular compartment comprising introducing a
XX DNA sequence encoding a heavy chain immunoglobulin, where the DNA is
XX linked to promoters and provided with an additional sequence encoding a
XX peptide capable of targeting heavy chain immunoglobulin to a cellular
XX compartment. The method is used for producing a heavy chain
XX immunoglobulin or an active fragment or derivative, or a protein that is
XX functionally equivalent for increasing the pathogen resistance in a plant
XX or to modulate metabolism in a plant. Under some circumstances it may be
XX desirable to retain the antibody product with the plant rather than
XX extracting and isolating the product. In particular, edible selected
XX CC antigens may be used in a method of passively immunising an animal.
XX CC preferably human, against the antigen, e.g., pathogenic organisms. The
XX CC present sequence is an anti-potato SBEII antibody (denoted Clone46) heavy
XX chain variable domain (VH) attached to peptide linkers, myc and his6 tag.
XX The potato SBEII is a starch branching enzyme also designated SBE A
XX
XX Sequence 152 AA;
XX
XX Query Match 71.6%; Score 489; DB 4; Length 152;
XX Best Local Similarity 74.6%; Pred. No. 2e-36;
XX Matches 100; Conservative 6; Mismatches 16; Indels 12; Gaps 2;
XX
XX 1 QVQLQESGGGLVQAGGSLRLSCAASGSIHTMGWYRQTPGTERDVAATIDGGSTNYA 60
XX 1 QVQLQESGGGLVQAGGSLRLSCAASGNTFSITAMAYRQAPGQRVAVSINSIGSTNYA 60
XX
XX 61 DSVKGRFTISRDNIINTVYLQWNSLKPEDTAIVHCHAD----VPRYTRRYIELMGQGLT 116
XX 61 DSVKGRFTISRDNAKNTVYLQWNSLKPEDTAIVYCAANGLVVRPY-----WGQGLT 112
XX
XX 117 VTVSSEPKTKPKPOP 130
XX 113 VTVSSEPKTKPKPOP 126
XX
XX
XX RESULT 9
XX AAE10563
XX ID AAE10563 standard; peptide; 124 AA.
XX
XX AAE10563;
XX
XX 10-DEC-2001 (first entry)
XX

```

	DE	HGL inhibiting VHH fragment; HGL #9 from llama species.
KW	Llama antibody; camelid; anorectic; heavy chain variable domain; VHH;	
KM	human dietary enzyme inhibitor; medicament; human pancreatic lipase; HPL;	
KX	food; human gastric lipase; HGL; cosmetic control; body weight.	
XX		
OS	Lama sp.	
XX		
FH	Key	
FT	Region	
FT	/label= CDR1	
FT	/note="Complementarity determining region 1"	
FT	Region	
FT	/label= CDR2	
FT	/note="Complementarity determining region 2"	
FT	Region	
FT	/label= CDR3	
FT	/note="Complementarity determining region 3"	
NN	EPI134231-AI,	
XX		
PD	19-SEP-2001.	
XX		
PP	20-FEB-2001; 200IEP-00200703.	
XX		
PR	14-MAR-2000; 200OEP-00200930.	
PA	(UNIL ) UNILEVER NV.	
PA	(UNIL ) UNILEVER PLC.	
XI	Besemer S., Van De Burg M., De Haard JW, Tareilus E;	
XX		
DX	WPt; 2001-572718/65.	
XX		
PT	New antibody or its fragments for inhibiting human dietary enzymes,	
PT	useful for cosmetic control of body weight of human beings, comprises	
PT	heavy chain variable domain derived from immunoglobulin naturally devoid	
PT	of light chains.	
PS		
ES	Example 4 ; Page 13; 37pp; English.	
XX		
CC	The patent discloses antibodies or their fragments comprising a heavy	
CC	chain variable domain (VHH) derived from an immunoglobulin naturally	
CC	devoid of light chains specific for inhibiting human dietary enzymes. The	
CC	antibodies of the invention are useful for the preparation of medicaments	
CC	or food for inhibiting the activity of one or more human dietary enzymes	
CC	especially human pancreatic lipase (HPL) or human gastric lipase (HGL)	
CC	which are useful for the cosmetic control of body weight of human beings.	
CC	The present peptide sequence is HGL inhibiting VHH fragment, HGL #9 from	
CC	llama (camelid) species	
XX		
SQ	Sequence 124 AA.	
OY	Query Match                71.4%; Score 488; DB 4; Length 124; Best Local Similarity 73.8%; Pred. No. 1.9e-36; Matches 96; Conservative 7; Mismatches 21; Indels 6; Gaps 1	
Db	1 QVOLOSGGVLVAGGSILLSCAASGSTIGSIHTMKWRQTGTGRDVATIQGGSTNYA 60 :.:.:.:.:.:.:.:.:.:.::	
OY	DSVKGFETSRNDLIITLVYLQNNSLKPEDTAVHYCNADVPYPRTSRLLEIMOGSLVTYS 120 :.:.:.:.:.:.:.:.:.:.::	
Dd	61 DSVKGFETISRNDIKXTMTWLQNNSLKPEDTGYYICAG-----TGAHGHWGGCTGYTVS 114 :.:.:.:.:.:.:.:.:.:.::	
OY	121 SEPKTPPOP 130 :.:.:.:.:.:.:.:.:.:.::	
Dd	115 SEPKEPKQP 124 :.:.:.:.:.:.:.:.:.:.::	

ABG30620;  
21-OCT-2002 (first entry)  
Immunoglobulin G specific heavy chain variable domain antibody #3.  
Heavy chain variable domain; antibody; protein array; aging; VHH;  
immunoglobulin; sticky phage-antibody; micro-panning; immune library;  
proteomic; mouse.  
Mus sp.  
WO200248193-A2.  
20-JUN-2002.  
03-DEC-2001; 2001WO-EP014471.  
13-DEC-2000; 2000EP-00311142.  
(UNIL ) UNILEVER NV.  
(UNIL ) UNILEVER PLC.  
(UNIL ) HINDUSTAN LEVER LTD.  
De Haard JWM, Hermans P, Landa I, Verrips CT;  
WPI; 2002-583487/62.  
Novel protein array useful for detecting the presence of individual  
proteins in sample, comprises heavy-chain variable domain antibodies or  
antibody fragments obtainable from Camelidae.  
Disclosure; Fig 3; 80bp; English.  
The invention describes a protein array (I) comprising a number of heavy-  
chain variable domain antibodies or antibody fragments, obtainable from  
Camelidae. The method is useful for removing abundant proteins from an  
extract or sample which do not provide useful information on the  
condition of a cell or tissue in the extract or sample to be  
investigated. (I) is useful for detecting the presence of individual  
proteins in a sample, comparing the distribution of proteins in different  
cell types, and identification of proteins that may be of importance in  
determining the altered properties of cells in disease, aging or other  
conditions. Using a heavy-chain variable domain derived from an  
immunoglobulin that is naturally devoid of light chains (VHH) in (I)  
provides a number of advantages, such as an improvement of  
sensitivity/resolution in the order of 10-100 times, and detection of  
post-translationally modified proteins. The invention also describes a  
method (II) that enables the simultaneous processing of large numbers of  
target antigens in a controlled way. The incorporated phage-ELISA  
generates on-line information about the success or failure of a certain  
panning condition. This feature combined with the microtiter plate format  
allows the complete automation of the technology, based on computer-made  
decisions on the values of the phage-ELISA for continuation of a limited  
number of selections. In (II), due to the fact that many different  
conditions can be tested, varying amounts of input-phages can be used  
simultaneously in order to decrease the enrichment of sticky phage-  
antibodies. Micro-panning is an effective tool for selecting both naive,  
synthetic and immune libraries on large numbers of different target  
molecules, enabling the generation of large panels of antibodies in short  
time frames needed for the generation of arrays (proteomics). This  
sequence represents an immunoglobulin (IgG) heavy chain variable domain  
antibody (VHH)  
Sequence 131 AA;  
Query Match 70.4%; Score 480.5; DB 5; Length 131;  
Best Local Similarity 72.5%; Pred. No. 9.8e-36;  
Matches 95; Conservative 10; Mismatches 25; Indels 1; Gaps 1;

Db 1 QVQLQESGGGLVQPGGSLRLSCAASGSIIFGAVGVMHQAQPKRELVARITDGSNTYA 60  
QY 61 DSVKGRFTISDNINITYLQMSLKPEDTAVYHCNADVPYRTSRL-ELMGGGTLVTY 119  
Db 61 DSVKGRFTISDNINITYLQMSLKPEDTAVYHCNADVPYRTSRL-ELMGGGTLVTY 120  
QY 120 SEPKTPKPOP 130  
Db 121 SEPKTPKPOP 131

RESULT 11  
AAE10566  
ID AAE10566 standard; peptide; 124 AA.

AC AAE10566;  
DT 10-DEC-2001 (first entry)  
DE HGL inhibiting VHH fragment, HGL #15 from llama species.  
KW llama antibody; camelid; anorectic; heavy chain variable domain; VHH;  
human dietary enzyme inhibitor; medicament; human pancreatic lipase; HPL;  
food; human gastric lipase; HGL; cosmetic control; body weight.  
XX Lama sp.

PH Key Location/Qualifiers  
FT 30..35  
FT /label= CDR1  
FT /note= "Complementarity determining region 1"  
FT Region 50..64  
FT /label= CDR2  
FT /note= "Complementarity determining region 2"  
FT 98..104  
FT /label= CDR3  
FT /note= "Complementarity determining region 3"

XX EP1134231-A1.  
XX 19-SEP-2001.  
XX PD  
XX PF 20-FEB-2001; 2001EP-00200703.  
XX PR 14-MAR-2000; 2000EP-00200930.  
XX PA (UNITL ) UNILEVER NV.  
XX PA (UNITL ) UNILEVER PLC.  
XX PI Bezemer S, Van De Burg M, De Haard JWM, Tareilus E;  
XX WPI; 2001-572718/65.

DR WPI; 2001-572718/65.  
XX New antibody or its fragments for inhibiting human dietary enzymes,  
PT useful for cosmetic control of body weight of human beings, comprises  
PT heavy chain variable domain derived from immunoglobulin naturally devoid  
PT of light chains.  
XX Example 4; Page 14; 37pp; English.

PS The patent discloses antibodies or their fragments comprising a heavy  
XX chain variable domain (VHH) derived from an immunoglobulin naturally  
CC devoid of light chains specific for inhibiting human dietary enzymes. The  
CC antibodies of the invention are useful for the preparation of medicaments  
CC or food for inhibiting the activity of one or more human dietary enzymes  
CC especially human pancreatic lipase (HPL) or human gastric lipase (HGL)  
CC which are useful for the cosmetic control of body weight of human beings.  
CC The present peptide sequence is HGL inhibiting VHH fragment, HGL #15 from  
CC llama (camelid) species  
XX Sequence 124 AA;

Query Match 69.7%; Score 476; DB 4; Length 124;

Best Local Similarity 72.3%; Pred. No. 2.4e-35;  
Matches 94; Conservative 7; Mismatches 23; Indels 6; Gaps 1;

QY 1 QVQLQESGGGLVQAGSGLRSLSCAASGSIISHTMGWYRTPTGTERDVAITDQGSNTYA 60  
Db 1 QVQLQESGGGLVQAGSGLRSLSCAASGSIISHTMGWYRTPTGTERDVAITDQGSNTYA 60  
QY 61 DSVKGRFTISDNINITYLQMSLKPEDTAVYHCNADVPYRTSRL-ELMGGGTLVTY 120  
Db 61 DSVKGRFTISDNINITYLQMSLKPEDTAVYHCNADVPYRTSRL-ELMGGGTLVTY 114

RESULT 12  
AAE10558  
ID AAE10558 standard; peptide; 129 AA.

AC AAE10558;  
DT 10-DEC-2001 (first entry)  
DE HPL inhibiting VHH fragment, HPL #22 from llama species.  
KW llama antibody; camelid; anorectic; heavy chain variable domain; VHH;  
human dietary enzyme inhibitor; medicament; human pancreatic lipase; HPL;  
food; human gastric lipase; HGL; cosmetic control; body weight.  
XX Lama sp.

PH Key Location/Qualifiers  
FT 31..35  
FT /label= CDR1  
FT /note= "Complementarity determining region 1"  
FT Region 50..64  
FT /label= CDR2  
FT /note= "Complementarity determining region 2"  
FT 98..109  
FT /label= CDR3  
FT /note= "Complementarity determining region 3"

XX EP1134231-A1.  
XX 19-SEP-2001.  
XX PD  
XX PF 20-FEB-2001; 2001EP-00200703.  
XX PR 14-MAR-2000; 2000EP-00200930.  
XX PA (UNITL ) UNILEVER NV.  
XX PA (UNITL ) UNILEVER PLC.  
XX PI Bezemer S, Van De Burg M, De Haard JWM, Tareilus E;  
XX WPI; 2001-572718/65.

DR WPI; 2001-572718/65.  
XX New antibody or its fragments for inhibiting human dietary enzymes,  
PT useful for cosmetic control of body weight of human beings, comprises  
PT heavy chain variable domain derived from immunoglobulin naturally devoid  
PT of light chains.  
XX Example 2; Page 10; 37pp; English.

PS The patent discloses antibodies or their fragments comprising a heavy  
XX chain variable domain (VHH) derived from an immunoglobulin naturally  
CC devoid of light chains specific for inhibiting human dietary enzymes. The  
CC antibodies of the invention are useful for the preparation of medicaments  
CC or food for inhibiting the activity of one or more human dietary enzymes  
CC especially human pancreatic lipase (HPL) or human gastric lipase (HGL)  
CC which are useful for the cosmetic control of body weight of human beings.  
CC The present peptide sequence is HPL inhibiting VHH fragment, HPL #22 from

CC llama (camelid) species  
 XX  
 SO Sequence 129 AA;  
 Query Match 68.4%; Score 467.5; DB 4; Length 129;  
 Best Local Similarity 72.3%; Pred. No. 1.4e-34;  
 Matches 94; Conservative 11; Mismatches 24; Indels 1; Gaps 1;

QY 1 QVQLQESGGGLVQAGSLRLSCAASGSISSIMTGMVYRQTPGERVAVATIQDGSITNYA 60  
 DB 1 QVQLQESGGGLVQAGSLRLSCAASGSISSIMTGMVYRQTPGERVAVATIQDGSITNYA 60  
 QY 61 DSVKGRFTISRDNILNTVYLQNNSLKPEPTAVYHCNADVPRTSRYLELMQGGTLVTVS 120  
 DB 61 DSVKGRFTISRDNILNTVYLQNNSLKPEPTAVYHCNADVPRTSRYLELMQGGTLVTVS 119  
 QY 121 SEPTKPKPOP 130  
 DB 120 SEPTKPKPOP 129

RESULT 13  
 AAE05288  
 ID AAE05288 standard; protein; 149 AA.  
 XX  
 AC AAE05288;  
 XX  
 DT 18-SEP-2001 (first entry)  
 XX  
 DE Anti-potato SBEII (Clone68) VH region attached with myc and his6 tag.  
 XX  
 KM Potato; heavy chain immunoglobulin; pathogen resistance;  
 KM metabolism modulator; passive immunisation; heavy chain variable domain;  
 KM VH; anti-potato SBEII; starch branching enzyme; SBE A.  
 XX  
 OS Solanum tuberosum.  
 OS Unidentified.  
 OS Chimeric.  
 XX  
 PN EP118669-A2.  
 XX  
 PD 25-JUL-2001.  
 XX  
 PF 08-DEC-2000; 2000EP-00310997.  
 XX  
 PR 17-DEC-1999; 99EP-00310188.  
 XX  
 PA (UNITL ) UNILEVER PLC.  
 PA (UNITL ) UNILEVER NV.  
 XX  
 PI Frenken LGJ, Van Der Logt CPE, Jobling SA, Teh Y;  
 DR WPI; 2001-427157/46.  
 DR N-PSDB; AADI0059.  
 XX  
 PT Modifying a plant to produce an antibody useful for increasing pathogen  
 PT resistance or to modulate metabolism comprises introducing a DNA sequence  
 PT encoding a heavy chain immunoglobulin linked to a peptide that targets a  
 PT cellular compartment.  
 XX  
 PS Example 13; Fig 28; 81pp; English.  
 XX  
 CC The present invention relates to a method for modifying a plant to  
 CC produce an antibody or an active fragment or derivative, or a protein  
 CC functionally equivalent, in a cellular compartment comprising introducing a  
 CC DNA sequence encoding a heavy chain immunoglobulin, where the DNA is  
 CC linked to promoters and provided with an additional sequence encoding a  
 CC peptide capable of targeting heavy chain immunoglobulin to a cellular  
 CC compartment. The method is used for producing a heavy chain  
 CC immunoglobulin or an active fragment or derivative, or a protein that is  
 CC functionally equivalent for increasing the pathogen resistance in a plant  
 CC or to modulate metabolism in a plant. Under some circumstances it may be  
 CC desirable to retain the antibody product with the plant rather than

CC extracting and isolating the product. In particular, edible selected  
 CC antigens may be used in a method of passively immunising an animal,  
 CC preferably human, against the antigen, e.g., pathogenic organisms. The  
 CC present sequence is an anti-potato SBEII antibody (denoted Clone68) heavy  
 CC chain variable domain (VH) attached to peptide linkers, myc and his6 tag.  
 CC The potato SBEII is a starch branching enzyme also designated SBE A  
 XX  
 SO Sequence 149 AA;  
 Query Match 68.3%; Score 466.5; DB 4; Length 149;  
 Best Local Similarity 73.1%; Pred. No. 2.1e-34;  
 Matches 95; Conservative 6; Mismatches 20; Indels 9; Gaps 1;

QY 1 QVQLQESGGGLVQAGSLRLSCAASGSISSIMTGMVYRQTPGERVAVATIQDGSITNYA 60  
 DB 3 QVQLQESGGGLVQAGSLRLSCAASGSISSIMTGMVYRQTPGERVAVATIQDGSITNYA 62  
 QY 61 DSVKGRFTISRDNILNTVYLQNNSLKPEPTAVYHCNADVPRTSRYLELMQGGTLVTVS 120  
 DB 63 DSVKGRFTISRDNILNTVYLQNNSLKPEPTAVYHCNADVPRTSRYLELMQGGTLVTVS 113  
 QY 121 SEPTKPKPOP 130  
 DB 114 SEPTKPKPOP 123

RESULT 14  
 AAE10561  
 ID AAE10561 standard; peptide; 129 AA.  
 XX  
 AC AAE10561;  
 XX  
 DT 10-DEC-2001 (first entry)  
 XX  
 DE HGL inhibiting VH fragment, HGL #4 from llama species.  
 XX  
 KM llama antibody; camelid; anorectic; heavy chain variable domain; VH;  
 KM human dietary enzyme inhibitor; medicament; human pancreatic lipase; HPL;  
 KM food; human gastric lipase; HGL; cosmetic control; body weight.  
 XX  
 OS Lama sp.  
 XX  
 FH Key  
 FH Region  
 FT Location/Qualifiers  
 FT 31..35  
 FT /label=CDR1  
 FT /note="Complementarity determining region 1"  
 FT Region  
 FT 50..64  
 FT /label=CDR2  
 FT /note="Complementarity determining region 2"  
 FT Region  
 FT 98..109  
 FT /label=CDR3  
 FT /note="Complementarity determining region 3"  
 XX  
 PN EP1134231-A1.  
 XX  
 PD 19-SEP-2001.  
 XX  
 PF 20-FEB-2001; 2001EP-00200703.  
 XX  
 PR 14-MAR-2000; 2000EP-00200930.  
 XX  
 PA (UNITL ) UNILEVER NV.  
 PA (UNITL ) UNILEVER PLC.  
 XX  
 PI Besemer S, Van De Burg M, De Haard JTW, Tareilus E;  
 DR WPI; 2001-572718/65.  
 XX  
 CC New antibody or its fragments for inhibiting human dietary enzymes,  
 CC useful for cosmetic control of body weight of human beings, comprises  
 CC heavy chain variable domain derived from immunoglobulin naturally devoid  
 CC of light chains.  
 PT

PS Example 4; Page 13; 37pp; English.

CC The patent discloses antibodies or their fragments comprising a heavy chain variable domain (VH) derived from an immunoglobulin naturally devoid of light chains specific for inhibiting human dietary enzymes. The antibodies of the invention are useful for the preparation of medicaments or food for inhibiting the activity of one or more human dietary enzymes especially human pancreatic lipase (HPL) or human gastric lipase (HGL) which are useful for the cosmetic control of body weight of human beings. The present peptide sequence is HGL inhibiting VH fragment, HGL #4 from llama (camelid) species

CC llama (camelid) species

CC Sequence 129 AA;

CC SQ

Query Match 68.0%; Score 464.5; DB 4; Length 129;  
Best Local Similarity 70.5%; Pred. No. 2.7e-34;  
Matches 93; Conservative 11; Mismatches 23; Indels 5; Gaps 2;

OY 1 QVQLQESGGGLVQAGGSLRLSCAASGSIQHTMGWYRQTPGERDVVATIQDGGSTNYA 60  
DB 1 QVQLQESGGGLVQAGGSLRLSCAASGSIQHTMGWYRQTPGERDVVATIQDGGSTNYA 60  
OY 61 DSVKGRFTISRDNTLTNTVYLQWNSLKPEDTAVYHC--NADVPYTSRYLEMGQGLT 118  
DB 61 DSVKGRFTISRDNTLTNTVYLQWNSLKPEDTAVYHC--NADVPYTSRYLEMGQGLT 118  
OY 119 VSSEPTKPKPOP 130  
DB 118 VSSEPTKPKPOP 129

RESULT 15  
AAE10567  
ID AAE10567 standard; peptide; 130 AA.  
AC AAE10567;  
XX 10-DEC-2001 (first entry)  
DT  
DE HGL inhibiting VH fragment, HGL #16 from llama species.  
XX  
XX llama antibody; camelid; anorectic; heavy chain variable domain; VH;  
KW human dietary enzyme inhibitor; medicament; human pancreatic lipase; HPL;  
KW food; human gastric lipase; HGL; cosmetic control; body weight.  
XX  
OS llama sp.  
XX  
FH Key Location/Qualifiers  
FH 31..35  
FT /label= CDR1  
FT /note= "Complementarity determining region 1"  
FT 50..64  
FT /label= CDR2  
FT /note= "Complementarity determining region 2"  
FT 98..110  
FT /label= CDR3  
FT /note= "Complementarity determining region 3"  
XX  
XX EP1134231-A1.  
XX PD 19-SEP-2001.  
XX PF 20-FEB-2001; 2001EP-00200703.  
XX PR 14-MAR-2000; 2000EP-00200930.  
XX PA (UNITL ) UNILEVER NV.  
XX PA (UNITL ) UNILEVER PLC.  
XX PI Bezemer S, Van De Burg M, De Haard JW, Tareilus E;  
XX WPI; 2001-572718/65.

PT New antibody or its fragments for inhibiting human dietary enzymes,  
PT useful for cosmetic control of body weight of human beings, comprises  
PT heavy chain variable domain derived from immunoglobulin naturally devoid  
PT of light chains.

CC Example 4; Page 14; 37pp; English.

CC The patent discloses antibodies or their fragments comprising a heavy chain variable domain (VH) derived from an immunoglobulin naturally devoid of light chains specific for inhibiting human dietary enzymes. The antibodies of the invention are useful for the preparation of medicaments or food for inhibiting the activity of one or more human dietary enzymes especially human pancreatic lipase (HPL) or human gastric lipase (HGL) which are useful for the cosmetic control of body weight of human beings. The present peptide sequence is HGL inhibiting VH fragment, HGL #16 from llama (camelid) species

CC llama (camelid) species

CC SQ

Query Match 67.8%; Score 463; DB 4; Length 130;  
Best Local Similarity 71.5%; Pred. No. 3.7e-34;  
Matches 98; Conservative 5; Mismatches 20; Indels 14; Gaps 2;

OY 1 QVQLQESGGGLVQAGGSLRLSCAASGSIQHTMGWYRQTPGERDVVATIQDGGSTNYA 60  
DB 1 QVQLQESGGGLVQAGGSLRLSCAASGSIQHTMGWYRQTPGERDVVATIQDGGSTNYA 60  
OY 61 DSVKGRFTISRDNTLTNTVYLQWNSLKPEDTAVYHCNADVPYTSRYLE-----WQ 113  
DB 61 DSVKGRFTISRDNTLTNTVYLQWNSLKPEDTAVYHCNADVPYTSRYLE-----WQ 113  
OY 114 GTLVVSSEPTKPKPOP 130  
DB 114 GTLVVSSEPTKPKPOP 130

Search completed: October 5, 2004, 08:00:11  
Job time : 99.9043 secs

This page blank (uspto)



A:Title: Humanrheumatoid factor cross-[idiotypes](#). IV. Studies on WA Xid-positive Igm wab  
 A:First from the 17.109 and GE Xids  
 A:Reference number: 155673; MUID:94065558; PMID:8245772  
 A:Accession: 155673  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-121 <RES>  
 A:Cross-references: GB:M87268; NID:g16197; PIDN:AA037536.1; PID:g16198  
 C:Genetics:  
 A:Gene: GDB:IGHM  
 A:Cross-references: GDB:120086; OMIM:147020  
 A:Map position: 14q32.33-14q32.33  
 C:Superfamily: immunoglobulin V region; immunoglobulin homology  
 A:15-98/Domain: immunoglobulin homology <IMM>

Query Match	56.9%;	Score 391.5;	DB 2;	Length 121;
Best Local Similarity	68.3%;	Pred. No. 2.4e-25;		
Matches 84;	Conservative 9;	Mismatches 27;	Indels 3;	Gaps 2;

QY I QVQLDSGGGLVQAGGSLRLSCAASGSISSINMGFRAPRGKDELVASIT-SGSSTNY 53  
:  
DB 1 EVQLLESGGGLVQPGGSLRLSTCAASGFPTFTYGMNWVRQAPGKGLEWVSATISGSGSTYY 60

```

Oy      60  ADSLGRFTSRDNAKNAVYQMNNLKPEDTAVYCNAAHTAGSSNVTGVGWHGTXT 118
        |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      61  ADSVGRFTSRDNSKNTLYQMNSTPAEDTAVYCAAPRHAGSP--PYDYWGGLVLT 118

```

QY	120	VSS	122
Db	119	VSS	121

RESULT 3  
S31107  
IS born

Ig heavy chain human  
 C:Species: Homo sapiens (man)  
 C>Date: 02-Dec-1993 #sequence\_revision 26-May-1995 #text\_change 17-Mar-1999  
 C|Accession: S31107 |Timmers, E.; Kenter, M.J.H.; van Tol, M.J.D.; Voessen, J.M.; Schuurman  
 R.Raaphorst, F.W.;  
 Eur. J. Immunol. 22, 247-251, 1992  
 A>Title: Restricted utilization of germ-line V(H)3 genes and short diverse third comple  
 A|Reference number: S31104; MUID:92111633; PMID:1730252  
 A|Accession: S31107  
 A|Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A|Molecule type: mRNA  
 A|Residues: 1-119 <RAA>  
 A|Cross-references: EMBL:X62955  
 A|Note: The nucleotide sequence was submitted to the EMBL Data Library, October 1991  
 C|Superfamily: immunoglobulin V region; immunoglobulin homology  
 C|Keywords: heterotetramer; immunoglobulin  
 F:15-98|Domain: immunoglobulin homology <Imm>

Query Match	56.8%;	Score 390.5;	DB 2;	Length 119;
Best Local Similarity	65.9%;	Pred. No. 2.9e-25;		
Matches	81;	Conservative	12;	Mismatches 25;
				Indels 5;
				Gaps 2

Qy I QVQLVESGGGLVQAGGSLRLSCAASGSISSINVMGFPRAPKQRRLVASIT-SGSSTNY 59  
:  
Db 1 EVQLVESGGGLVQPGGSLRLSCAASGFTFSSYAMSWVRQAPKGLEWVSATISGSGSTYY 60

```

Qy      60 ADSLGRFTISRDNAKNAVYLQNNUKPEDPAVYVYCNAAHTIPAGSSNYVYGVWGHGKVT 119
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      61 ADSVGRFTISRDNKQNTLYLQNNSLRAEDPAVYVC---AKDPGASYYPDYWGQGTLLVT 116

```

QY	120	VSS	122
Db	117	VSS	119

RESULT 4  
S31588  
Ig heavy chain V region - human (fragment)  
C;Species: Homo sapiens (man)

C, Date: 22-Nov-1993 #sequence\_revision 10-Nov-1995 #text\_change 23-Jul-1999  
C, Accession: S31588  
R, Cuisinier, A.M.; Gauthier, L.; Boubli, L.; Fougereau, M.; Tonnelie, C.  
submitted to the EMBL Data Library, June 1992  
A, Description: Mechanisms that generate human immunoglobulin diversity operate from the  
A, Reference number: S31585  
A, Accession: S31588  
A, Status: preliminary  
A, Molecule type: mRNA  
A, Residues: 1-140 <CUI>  
A, Cross-references: EMBL:Z14200; NID:g30957; PIDN:CAA78566.1; PID:g30958  
C, Superfamily: immunoglobulin V region; immunoglobulin homology  
C, Keywords: heterotetramer; immunoglobulin  
F, 34-117/Domain: immunoglobulin homology <IMM>

Query March	56.5%	Score 388.5;	DB 2;	Length 140;
Best Local Similarity	66.9%	Pred. No. 4.9e-25;		
Matches 83; Conservative	11;	Mismatches 25;	Indels 5;	Gaps 3

DQ  
Db

1 QVQVQESGGGLVQAGGSLRLISCAASSISSINVMGWRCQAPGRKELVASIT-EGGSINI 50  
:  
: ||| | | | | | | | | : : :  
20 EVQLLESGGGLVQPGGSLRLISCAASFTFSYAMSWVRQAQPGGLEWYSALISGGGSTYY 79

Dy 60 ADSLKGKFTISRDKNAAYLYQNNLRPESTAVYCNALHPAGSSVYI-VSWGHGIV IIG  
|||:|||::||:||||:||||| |||: ||| |||  
Db 80 ADSVKGRFTISRDSKNLTLYQNMSLRAPETAVYC--AKDHDYSNYIYPDYWGQGTLY 136

QY	119	TVSS	122
Db	137	TVSS	140

RESULT 5  
S31110  
Id heavy

ig heavy chain - human  
C:Species: Homo sapiens (man)  
C:Date: 02-Dec-1993 #sequence\_revision 26-May-1995 #ext\_change 17-Mar-1999  
C:Accession: S3110  
R:Raaphorst, F.M.; Timmers, E.; Kenter, M.J.H.; van Tol, M.J.D.; Vossen, J.M.; Schuurman  
E:Utr.: J. Immunol.: 22, 247-251, 1992  
A:Title: Restricted utilization of germ-line V(H)3 genes and short diverse third complem  
A:Reference number: S31104; M0ID:9211633; PMID:131052  
A:Accession: S3110  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: mRNA  
A:Residues: 1-116 <RAA>  
A:Cross-references: EMBL:X62958  
A:Note: The nucleotide sequence was submitted to the EMBL Data Library, October 1991  
C:Superfamily: immunoglobulin V region; immunoglobulin homology  
C:Keywords: heterotetramer; immunoglobulin  
F:15-97/Domain: immunoglobulin homology <IMM>

Query Match	56.4%	Score 388	DB 2	Length 116
Best Local Similarity	64.8%	Pred. No. 4.5e+25		
Matches 79	Conservative 11	Mismatches 26	Indels 6	Gaps 1

QY 1 QVQIQESGGGILVQAGSLRLSCAASGSISSINVMGWFRQAPGKRELVASITSGGSINVA 60  
: : : : : : : : : : : : : : : : : :  
DB 1 EVQLVETGGGLIQPGSLRLSCAASGVTVSSNIMSNVRQAPEGLEWVSIVTISGGSTYYA 60

D5  
QY

61 DSLKGRFTISRDNADAVYQMNNLKEEDIAVYYCAHITPAGSSNIVYGIMWGHGIRIV 120  
||:||:||:||::|::|::|::|::|::|::|  
D5 DSVTGRFITSRDNSKNLTLYRMSLRPAEDIAVIYCARDLGGGGD-----YMCGTLVTIY 118

QY	121	SS	122
Db	115	SS	116

RESULT 6  
S3114  
Ig heavy chain - human  
C/Species: Homo sapiens (man)



C>Date: 02-Dec-1993 #sequence\_revision 26-May-1995 #text\_change 17-Mar-1999  
C:Accession: S31114  
R:Raaphorst, P.M.; Timmers, E.; Kenter, M.J.H.; van Tol, M.J.D.; Vossen, J.M.; Schuurman  
Eur. J. Immunol. 22, 247-251, 1992  
A>Title: Restricted utilization of germ-line V(H)3 genes and short diverse third comple  
A:Accession number: S31104; MUID:92111633; PMID:1730252  
A:Accession: S31114  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: mRNA  
A:Residues: 1-123 <RAA>  
A:Cross-references: EMBL:X62963  
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1991  
C:Superfamily: immunoglobulin V region; immunoglobulin homology  
C:Keywords: heterotetramer; immunoglobulin  
F:15-98/Domain: immunoglobulin homology <IMM>

Query Match	56.0%	Score 385.5;	DB 2;	Length 123;
Best Local Similarity	64.3%;	Pred. No. 7.6e-25;		
Matches	83;	Conservative	12;	Mismatches 21;
			Indels	13;
			Gaps	3;

Dy 1 QVQLDGGGGLVQAGSLLRLSCAASGISINMGMFRQAPKQRELVASIT-  
SGGSINY 59  
:  
Db 1 EVQLIESGGGLVQPGGSLRLSCAASGETFSSYAMSWVRQAPKGLEWVSATISGGSTYY 60

```
QY      60 ADSLGRFTISRDNKNAVYLQMNNTKPEDTAVVYCNAAHTTPAGSSNY-----VGYWG 113
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      61 ADVSKGRFTISRDNKNTLYLQMSNRADPTAVVYCY-----AKASLYLRPLEWLPDYWG 114
```

QY	114	HGRTVTSS	122
Db	115	QGLVTSS	123

## RESULT 7

I9 heavy chain human  
C:Species: Homo sapiens (man)  
C:Date: 02-Dec-1993 #sequence\_revision 26-May-1995 #text\_change 17-Mar-1999  
C:Accession: S31108  
R:Rapoport, F.M., Timmers, E., Kenter, M.J.H., van Tol, M.J.D., Vossen, J.M., Schuurman  
Eur. J. Immunol. 22, 247-251, 1992  
A>Title: Restricted utilization of germ-line V(H)3 genes and short diverse third comple  
A:Reference number: S31104; MIMD:92111633; EMID:1730252  
A:Accession: S31108  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: mRNA  
A:Residues: 1-119 <RAA>  
A:Cross-references: EMBL:X62956  
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1991  
C:Superfamily: immunoglobulin V region; immunoglobulin homology  
C:Keywords: heterotetramer; immunoglobulin  
/15-98/Domain: immunoglobulin homology <IMM>

Query Match	55.7%;	Score 383.5;	DB 2;	Length 119;
Best Local Similarity	65.9%;	Pred. No. 1.1e-24;		
Matches	81;	Conservative	11;	Mismatches 26;
			Indels	5;
			Gaps	2;

QY 1 QVQLQESGGGLVQAGGSLRLSCAAGSISINMGVFRQAPGKOREIIVASTT-SGGSTNY 59  
:|||||:|||||:|:|:|  
Db 1 EVQLLESGGGLVQPGGSLRLSCAAGSFTFSSYAMSVNRQAPGKGLEWVAISGSGSTYY 60

QY : 60 ADSLKGFRFISRDNAKANVYLQMNLIKPEDTAVYYCAHITPAGSSNYYVGWGHGKVT 119  
|||:|||||:|||||:|||||:|||||:  
Dd 61 ADSVKGRFISRDNSKNLTLYIQMNSLRAPEDTAVYYCAKDRLTG-----FDWGQGTLVT 116

Qy	120	VSS	122
Db	117	VSS	119

RESULT 8  
S31686  
Ig heavy chain V region - human (fragment)

```

C:Species: Homo sapiens (man)
C:Date: 22-Nov-1993 #sequence_revision 10-Nov-1995 #next_change 23-Jul-1999
C:Accession: S31586
R:Cusinier, A.M.; Gauthier, L.; Boubli, L.; Fougereau, M.; Tonnelle, C.
Submitted to the EMBL Data Library, June 1992
A:Description: Mechanisms that generate human immunoglobulin diversity operate from the
A:Reference number: S31585
A:Accession: S31586
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-140 <CUI>
A:Cross-references: EMBL:Z14205; NID:G30969; PIDN:CAA78574.1; PID:G30970
C:Superfamily: immunoglobulin V region; immunoglobulin homology
C:Keywords: heterotetramer; immunoglobulin
;34-117/Domain: immunoglobulin homology <IMM>

```

Query Match	55.7%	Score 383.5	DB 2	Length 140
Best Local Similarity	65.9%	Pred. No. 1.3e-24		
Matches 81	Conservative 12	Mismatches 27	Indels 3	Gaps 2

QY 1 QYVLOEGGGGAYCAGGSLRLCSAAGSISINVMGWFRQAPKQRELVASIR - SGGSITNY 59

Db 20 EYVLLBEGGGVLPGGSLRLCSAGSTFTFSIAMSVMRQAPGGLGVNVSISGGSGSTYY 79

QY 60 AASLKSGFTISPDNAKAAVYOMNLKPEDTAVVYCNAAITPAGSSNVYVGGWGHGKVT 119

Db 80 SOSVKGFTISPDNSKNTLLYOMNSLAEEDTAVVYCAKCPFAGGSB- -FDWGGQGLVT 137

Qy	120	VSS	122
	11		
	11		
Db	138	VSS	140

## RESULT 9

Ig heavy chain V region (30pt) - human  
C:Species: Homo sapiens (man)  
C:Date: 21-Dec-1990 #sequence\_revision 21-Dec-1990 #text\_change 16-Aug-1996  
C:Accession: C36005  
R:Schroeder Jr., H.W.; Wang, J.Y.  
Proc. Natl. Acad. Sci. U.S.A. 87, 6146-6150, 1990  
A:Title: Preferential utilization of conserved immunoglobulin heavy chain variable gene  
A:Reference number: A36005; M0ID:90349571; PMID:2117273  
A:Accession: C36005  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-119 <SCH>  
A:Cross-references: GB:M18513  
C:Superfamily: immunoglobulin V region; immunoglobulin homology  
C:Keywords: heterotrimer; immunoglobulin  
F:15-98/Domain: immunoglobulin homology <IMM>

Best Local	Similarity	67.58	Pred. No. 1.3e-24	Indels	Gaps
Matches	83	Conservative	10	Mismatches	25

Db  
1 EVQLLEGGGGLVQPGASLRISCAASGFTFSYAMSNVROAPGKGLKEWVAISLGGSGSY 60

Db 61 ADSVKGKFTLISRDNSKNTLLYLOMNSLSAEDTAIVYIC-AKDAWGSG--FDIYGGGLVIT 116

Db	1.17	VSS	11.9
4	1.17	11.9	11.9
5	1.17	11.9	11.9
6	1.17	11.9	11.9
7	1.17	11.9	11.9
8	1.17	11.9	11.9
9	1.17	11.9	11.9
10	1.17	11.9	11.9
11	1.17	11.9	11.9
12	1.17	11.9	11.9
13	1.17	11.9	11.9
14	1.17	11.9	11.9
15	1.17	11.9	11.9
16	1.17	11.9	11.9
17	1.17	11.9	11.9
18	1.17	11.9	11.9
19	1.17	11.9	11.9
20	1.17	11.9	11.9
21	1.17	11.9	11.9
22	1.17	11.9	11.9
23	1.17	11.9	11.9
24	1.17	11.9	11.9
25	1.17	11.9	11.9
26	1.17	11.9	11.9
27	1.17	11.9	11.9
28	1.17	11.9	11.9
29	1.17	11.9	11.9
30	1.17	11.9	11.9
31	1.17	11.9	11.9
32	1.17	11.9	11.9
33	1.17	11.9	11.9
34	1.17	11.9	11.9
35	1.17	11.9	11.9
36	1.17	11.9	11.9
37	1.17	11.9	11.9
38	1.17	11.9	11.9
39	1.17	11.9	11.9
40	1.17	11.9	11.9
41	1.17	11.9	11.9
42	1.17	11.9	11.9
43	1.17	11.9	11.9
44	1.17	11.9	11.9
45	1.17	11.9	11.9
46	1.17	11.9	11.9
47	1.17	11.9	11.9
48	1.17	11.9	11.9
49	1.17	11.9	11.9
50	1.17	11.9	11.9
51	1.17	11.9	11.9
52	1.17	11.9	11.9
53	1.17	11.9	11.9
54	1.17	11.9	11.9
55	1.17	11.9	11.9
56	1.17	11.9	11.9
57	1.17	11.9	11.9
58	1.17	11.9	11.9
59	1.17	11.9	11.9
60	1.17	11.9	11.9
61	1.17	11.9	11.9
62	1.17	11.9	11.9
63	1.17	11.9	11.9
64	1.17	11.9	11.9
65	1.17	11.9	11.9
66	1.17	11.9	11.9
67	1.17	11.9	11.9
68	1.17	11.9	11.9
69	1.17	11.9	11.9
70	1.17	11.9	11.9
71	1.17	11.9	11.9
72	1.17	11.9	11.9
73	1.17	11.9	11.9
74	1.17	11.9	11.9
75	1.17	11.9	11.9
76	1.17	11.9	11.9
77	1.17	11.9	11.9
78	1.17	11.9	11.9
79	1.17	11.9	11.9
80	1.17	11.9	11.9
81	1.17	11.9	11.9
82	1.17	11.9	11.9
83	1.17	11.9	11.9
84	1.17	11.9	11.9
85	1.17	11.9	11.9
86	1.17	11.9	11.9
87	1.17	11.9	11.9
88	1.17	11.9	11.9
89	1.17	11.9	11.9
90	1.17	11.9	11.9
91	1.17	11.9	11.9
92	1.17	11.9	11.9
93	1.17	11.9	11.9

S31666  
Ig heavy chain V region - human (fragment)  
C;Species: Homo sapiens (man)





This page blank (uspio)

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2004, 07:29:35 ; Search time 11.7253 Seconds  
(without alignments)  
581.749 Million cell updates/sec

Title: US-09-805-290A-22

Perfect score: 688  
Sequence: 1 QVQLQESGGGLVQAGGSLRL.....MGHGTXTVTSSEPKTPKPP 131

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	349.5	50.8	115	1	HV3F_HUMAN
2	349	50.7	116	1	HV05_CARAU
3	347	50.4	126	1	HV3K_HUMAN
4	345	50.1	136	1	HV16_MOUSE
5	339.5	49.3	117	1	HV3C_HUMAN
6	338	49.1	114	1	HV3B_HUMAN
7	337	49.0	122	1	HV3G_HUMAN
8	335.5	48.8	119	1	HV3L_HUMAN
9	334	48.5	119	1	HV3P_HUMAN
10	333.5	48.5	115	1	HV3D_HUMAN
11	331.5	48.2	111	1	HV3T_MOUSE
12	330	48.0	116	1	HV3T_HUMAN
13	329.5	47.9	115	1	HV32_MOUSE
14	329.5	47.9	117	1	HV02_CANFA
15	327.5	47.6	113	1	HV30_MOUSE
16	326.5	47.5	119	1	HV3I_HUMAN
17	326	47.4	97	1	HV36_MOUSE
18	325	47.2	119	1	HV37_MOUSE
19	324	47.1	111	1	HV35_MOUSE
20	324	47.1	119	1	HV40_MOUSE
21	323	46.9	114	1	HV01_CANFA
22	323	46.9	119	1	HV38_MOUSE
23	323	46.9	119	1	HV3N_HUMAN
24	323	46.9	120	1	HV3U_HUMAN
25	323	46.9	122	1	HV3A_HUMAN
26	322.5	46.9	113	1	HV27_MOUSE
27	322.5	46.9	118	1	HV39_MOUSE
28	320.5	46.6	121	1	HV3J_HUMAN
29	319.5	46.4	120	1	HV3E_HUMAN
30	318.5	46.3	117	1	HV53_MOUSE
31	318	46.2	119	1	HV3K_HUMAN
32	317.5	46.1	117	1	HV42_MOUSE
33	316.5	46.0	113	1	HV28_MOUSE

34	315.5	45.9	142	1	HV01_RAT	P01805	rattus norv
35	314.5	45.7	117	1	HV55_MOUSE	P18526	mus musculu
36	314	45.6	122	1	HV21_MOUSE	P01790	mus musculu
37	312.5	45.4	98	1	HV57_MOUSE	P18528	mus musculu
38	312.5	45.4	113	1	HV29_MOUSE	P01790	mus musculu
39	312.5	45.4	113	1	HV31_MOUSE	P01800	mus musculu
40	312.5	45.4	117	1	HV54_MOUSE	P18525	mus musculu
41	312	45.3	122	1	HV3H_HUMAN	P01769	homo sapien
42	307.5	44.7	123	1	HV18_MOUSE	P01787	mus musculu
43	307	44.6	122	1	HV20_MOUSE	P01789	mus musculu
44	304.5	44.3	123	1	HV25_MOUSE	P01794	mus musculu
45	303.5	44.1	123	1	HV19_MOUSE	P01788	mus musculu

ALIGNMENTS

RESULT 1	HV3F_HUMAN	STANDARD;	PRT;	115 AA.
ID	HV3F_HUMAN			
AC	P01767			
DT	21-JUL-1986 (Rel. 01, Created)			
DT	21-JUL-1986 (Rel. 01, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	Ig heavy chain V-II region BUT.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
CX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE.			
RX	MEDLINE=78137069; PubMed=416441;			
RA	Torano A., Putnam F.W.;			
RT	"Complete amino acid sequence of the alpha 2 heavy chain of a human			
RT	Igk2 immunoglobulin of the A2m (2) allotype."			
RL	Proc. Natl. Acad. Sci. U.S.A. 75:966-969(1978).			
CC	MISCBLASTOBS: THE SEQUENCE OF THE ALPHA-2, A2M(2) ALLOTYPIC, C			
CC	-1- SIMILARITY: Contains 1 immunoglobulin-like domain.			
DR	HSSP; A02050; A2HUBU.			
DR	HSSP; P01789; IMCP.			
DR	GO; GO:0005576; C:extracellular; NAS.			
DR	GO; GO:0003823; F:antigen binding; NAS.			
DR	GO; GO:0006955; P:immune response; NAS.			
DR	InterPro; IPR007110; Ig-like.			
DR	InterPro; IPR003596; Ig_v.			
DR	Pfam; PF00047; Ig_1.			
DR	SMART; SM00406; IGV_1.			
DR	PROSITE; PS50835; IG_LIKE; 1.			
KW	Immunoglobulin V region.			
FT	DOMAIN 1 111 IG-LIKE.			
FT	NON_TER 115			
SEQUENCE	115 AA; 12379 MW; 208876A7DF52DCF4 CRC64;			
Query Match	50.8%; Score 349.5; DB 1; Length 115;			
Best Local Similarity	58.2%; Pred. No. 1.2e-26;			
Matches	71; Conservative 18; Mismatches 26; Indels 7; Gaps 1;			
QY	1 QVQLQESGGGLVQAGGSLRLSCAASGSISSINMGWFRQAPGKORELVASITSGGSTNYA 60			
DB	1 EVQLVETGGGLIQGGSIRLSCAASGFTVBHSMSTWRQAPGKALZWVSAIYRGSTIYYA 60			
QY	61 DSLKGRFTISRDNAKKNVYLQMNNLKPEDPAVYVCNANHITPAGSSNVYVYWGHTKVTY 120			
DB	61 DSVKGRFTISRDSDRBTIVYQMSLREPDPAVYVCARDLAA-----RLFGRKTTIYY 113			
QY	121 SS 122			
DB	114 SS 115			
RESULT 2				
HV05_CARAU				

```

ID      HV05_CARAU      STANDARD;      PRT;      116 AA.
AC      P19181;
DT      01-NOV-1990 (Rel. 16, Last sequence update)
DT      01-NOV-1990 (Rel. 16, Last sequence update)
DT      15-JUL-1999 (Rel. 38, Last annotation update)
DE      Ig heavy chain V region 5A precursor.
OS      Carassius auratus (Goldfish).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC      Cyprinidae; Carassius.
OX      NCBI_TaxID=7957;
RN      [1]
RP      SOURCE FROM N.A.
RX      MEDLINE=8144476; PubMed=3125551;
RA      Wilson M.R., Middleton D., Warr G.W.;
RT      "Immunoglobulin heavy chain variable region gene evolution: structure
RT      and family relationships of two genes and a pseudogene in a teleost
RT      fish."
RL      Proc. Natl. Acad. Sci. U.S.A. 85:1566-1570 (1988).
DR      HSP: P01772; 2PB4.
DR      InterPro: IPR007110; Ig-Like.
DR      InterPro: IPR003596; Ig_V.
DR      Pfam: PF00047; Ig; 1.
DR      SMART: SM00406; IgV; 1.
DR      PROSITE: PS50835; IG_LIKE; 1.
KW      Immunoglobulin V region; Signal.
FT      SIGNAL      1
FT      CHAIN      1
FT      DOMAIN      20 116      IG HEAVY CHAIN V REGION 5A.
FT      DOMAIN      20 49      FRAMEWORK-1.
FT      DOMAIN      50 54      COMPLEMENTARITY-DETERMINING-1.
FT      DOMAIN      55 69      FRAMEWORK-2.
FT      DOMAIN      69 84      COMPLEMENTARITY-DETERMINING-2.
FT      DOMAIN      85 116      FRAMEWORK-3.
FT      DISULFID      41 114      BY SIMILARITY.
FT      NON_TER      116
SQ      SEQUENCE      116 AA; 12808 MW; 9C2279E2DF199B12 CRC64;

Query Match      50.7%; Score 349; DB 1; Length 116;
Best Local Similarity 72.6%; Pred. No. 1.4e-26;
Matches 69; Conservative 8; Mismatches 18; Indels 0; Gaps 0;

QY      1 QVQLVSGGGLVQAGSLRLSCAASGSISSINVMGFRCAQPKRELVAISITSGG-STNYA 60
DB      20 EVQLVSSGGGLDLPQGSRLSCAASGFTVSNMVSVPQPKGLEWVSIVISGGSTYYA 79
QY      61 DSLKGRFTISRDNKAKAVYLQMNMLKPEPTAVYVC 95
DB      80 DSVKGRFTISRDNKSKTLYLQMNSLRAEDTAVYVC 114

RESULT 3
HV3K_HUMAN      STANDARD;      PRT;      126 AA.
ID      HV3K_HUMAN
AC      P01772;
DT      21-JUL-1986 (Rel. 01, Created)
DT      21-JUL-1986 (Rel. 01, Last sequence update)
DT      10-OCT-2003 (Rel. 42, Last annotation update)
DE      Ig heavy chain V-III region KOL.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX      NCBI_TaxID=9606;
RN      [1]
RP      SEQUENCE AND DISULFIDE BONDS.
RX      MEDLINE=83289131; PubMed=6884994;
RA      Schmidt W.E., Jung H.-D., Palm W., Hilschmann N.;
RT      "Three-dimensional structure determination of antibodies. Primary
RT      structure of crystallized monoclonal immunoglobulin IgG1 KOL, I.";
RL      Hoppe-Seyler's Z. Physiol. Chem. 364:713-747 (1983).
RN      [2]
RX      X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).
RX      MEDLINE=81072295; PubMed=7441755;
RA      Marguier M., Delsenhofer J., Huber R., Palm W.;

```

```

RT      "Crystallographic refinement and atomic models of the intact
RT      immunoglobulin molecule KOL and its antigen-binding fragment at 3.0 A
RT      and 1.0-A resolution."
RL      J. Mol. Biol. 141:369-391 (1980).
CC      -1 SIMILARITY: Contains 1 immunoglobulin-like domain.
DR      PIR: A02055; G1HXL.
DR      PDB: 2PB4; 12-JUL-89.
DR      PDB: 2IG2; 12-JUL-89.
DR      GO: GO:0005576; C:extracellular; NAS.
DR      GO: GO:0003823; F:antigen binding; NAS.
DR      GO: GO:0006955; P:immune response; NAS.
DR      InterPro: IPR007110; Ig-Like.
DR      InterPro: IPR003596; Ig_V.
DR      Pfam: PF00047; Ig; 1.
DR      SMART: SM00406; IgV; 1.
DR      PROSITE: PS50835; IG_LIKE; 1.
KW      Immunoglobulin V region; 3D-structure; Pyrrolidone carboxylic acid.
FT      DOMAIN      1
FT      MOD_RES      1 112      PYRROLIDONE CARBOXYLIC ACID.
FT      DISULFID      22 96
FT      DISULFID      105 110
FT      STRAND      3 7
FT      STRAND      11 12
FT      TURN      14 15
FT      STRAND      18 25
FT      STRAND      29 31
FT      HELIX      34 39
FT      STRAND      41 42
FT      TURN      45 51
FT      STRAND      53 54
FT      TURN      58 60
FT      STRAND      62 64
FT      HELIX      65 65
FT      STRAND      66 67
FT      TURN      68 73
FT      TURN      74 77
FT      STRAND      78 83
FT      STRAND      88 90
FT      HELIX      92 99
FT      STRAND      106 106
FT      STRAND      107 108
FT      TURN      109 109
FT      STRAND      113 116
FT      STRAND      120 124
FT      STRAND      126
FT      NON_TER      126
SQ      SEQUENCE      126 AA; 13718 MW; E4D71B52E16F8776 CRC64;

Query Match      50.4%; Score 347; DB 1; Length 126;
Best Local Similarity 59.5%; Pred. No. 2.3e-26;
Matches 75; Conservative 15; Mismatches 32; Indels 4; Gaps 2;

QY      1 QVQLVSGGGLVQAGSLRLSCAASGSISSINVMGFRCAQPKRELVAISITSGG-STNY 59
DB      1 QVQLVSSGGGLVQAGSLRLSCSSGFTFSVAMVWRQAPKGLWVAIIWDDGSDQHY 60
QY      60 ADSLKGRFTISRDNKAKAVYLQMNMLKPEPTAVYVC---NAHITPAGSSNYVYGVGHGT 116
DB      61 ADSVKGRTISRDNKSKTLYLQMNSLRPEPTGVFCARDGGHGFCSASCFGPDYWGQGT 120
QY      117 KVTYSS 122
DB      121 PVTYSS 126

RESULT 4
HV16_MOUSE      STANDARD;      PRT;      136 AA.
ID      HV16_MOUSE
AC      P01783;
DT      21-JUL-1986 (Rel. 01, Created)
DT      21-JUL-1986 (Rel. 01, Last sequence update)
DT      10-OCT-2003 (Rel. 42, Last annotation update)
DE      Ig heavy chain V region MOPC 21 precursor (Fragment).
OS      Mus musculus (Mouse).

```

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
CX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=81234548; PubMed=6788376;  
RA Botheil A.L.M., Paskind M., Reth M., Imanishi-Kari T., Rajewsky K.,  
BA Baltimore D.;  
RT "Heavy chain variable region contribution to the Npb family of  
RT antibodies: somatic mutation evident in a gamma 2a variable region";  
RL Cell 24:625-637(1981).  
CC [2]  
CC SEQUENCE OF 17-136.  
RP MEDLINE=77100368; PubMed=401950;  
RX Adetubo K., Milstein C., Secher D.S.;  
RA "Molecular analysis of spontaneous somatic mutants";  
RL Nature 265:299-304(1977).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; J00522; AAD15290.1; -  
DR PIR; E90809; GIMS21.  
DR PDB; 1IGC; 03-JUN-95.  
DR InterPro; IPR007110; IG-like.  
DR InterPro; IPR003596; IG\_v.  
DR Pfam; PF00047; Ig; 1.  
DR SMART; SMO0406; Igv; 1.  
DR PROSITE; PSS0835; IG LIKE; 1.  
KW Immunoglobulin V region; Signal; 3D-structure.  
FT SIGNAL 1  
FT CHAIN 16  
FT DOMAIN 17 136 IG HEAVY CHAIN V REGION MOPC 21.  
FT DOMAIN 115 119 D SEGMENT.  
FT DOMAIN 120 136 JH4 SEGMENT.  
FT DISULFID 38 112  
FT CONFLICT 75 78 HYAD -> DYAH (IN REF. 2).  
FT CONFLICT 89 90 DN -> ND (IN REF. 2).  
FT CONFLICT 115 115 W -> H (IN REF. 2).  
FT CONFLICT 120 120 Y -> W (IN REF. 2).  
FT NON\_TER 136  
SQ SEQUENCE 136 AA; 15071 MW; 2276A98DBDF7016 CRC64;  
Query Match 50.1%; Score 345; DB 1; Length 136;  
Best Local Similarity 61.6%; Pred. No. 3,9e-26;  
Matches 77; Conservative 10; Mismatches 28; Indels 10; Gaps 3;  
QY 2 VQLQSGGGLVQPGSLRISCAAGSISINVMGFRQAPGKQRLVASITSGGST- NYA 60  
DB 18 VQLVHSGGGLVQPGSGRLKSCAAGFTFSSFGHFWVRQAPKGLWVAIVISGGSSTLHYA 77  
QY 61 DSLKGRFTISRDNKAVYLVQNNLKPEDTAVYVYGNNAITAGSSNNYV- -GWGSGTK 117  
DB 78 DTVKGRFTISRDNKAVYLVQNNLKPEDTAVYVYGNNAITAGSSNNYV- -ARWGNPYVAMDTWGQTS 131  
QY 118 VTSSS 122  
DB 132 VTSSS 136  
RESULT 5  
HV3C\_HUMAN STANDARD; PRT; 117 AA.  
ID HV3C\_HUMAN  
AC P01764;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Ig heavy chain V-III region VH26 precursor.

OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
CX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=81101090; PubMed=6450418;  
RA Mathysens G., Rabbits T.H.;  
RT "Structure and multiplicity of genes for the human immunoglobulin  
RT heavy chain variable region";  
RL Proc. Natl. Acad. Sci. U.S.A. 77:6561-6565(1980).  
CC [1- SIMILARITY: Contains 1 immunoglobulin-like domain.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; J00236; AAA53516.1; -  
DR EMBL; M35415; AAA58735.1; -  
DR PIR; A02047; H3H26.  
DR PDB; 1H0U; 23-DEC-99.  
DR Genew; HGNC:5545; IGHV@.  
DR GO; GO:0005576; C:extracellular; NAS.  
DR GO; GO:0003823; P:antigen binding; NAS.  
DR GO; GO:0006955; P:immune response; NAS.  
DR InterPro; IPR007110; IG-like.  
DR InterPro; IPR003596; IG\_v.  
DR Pfam; PF00047; Ig; 1.  
DR SMART; SMO0406; Igv; 1.  
DR PROSITE; PSS0835; IG LIKE; 1.  
KW Immunoglobulin V region; Signal; 3D-structure.  
FT SIGNAL 1  
FT CHAIN 20 117 IG HEAVY CHAIN V-III REGION VH26.  
FT DOMAIN 20 117 IG-LIKE.  
FT NON\_TER 117  
SQ SEQUENCE 117 AA; 12582 MW; E826733F1A3CB0F1 CRC64;  
Query Match 49.3%; Score 339.5; DB 1; Length 117;  
Best Local Similarity 71.9%; Pred. No. 1.1e-25;  
Matches 69; Conservative 9; Mismatches 17; Indels 1; Gaps 1;  
QY 1 QVQLSGGGLVQPGSLRISCAAGSISINVMGFRQAPGKQRLVASITSGGSTVY 59  
DB 20 EVQLSGGGLVQPGSGRLKSCAAGFTFSSYAMSWVRQAPKGLWVAISGGSSTYY 79  
QY 60 ADSLKGRFTISRDNKAVYLVQNNLKPEDTAVYVY 95  
DB 80 GDSVKGRTISRDNKAVYLVQNNLKPEDTAVYVY 115  
RESULT 6  
HV3B\_HUMAN STANDARD; PRT; 114 AA.  
ID HV3B\_HUMAN  
AC P01763;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Ig heavy chain V-III region WEA.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
CX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE.  
RX MEDLINE=83273707; PubMed=6410398;  
RA Goni F., Frangione B.;  
RT "Amino acid sequence of the Fv region of a human monoclonal IgM  
RT (protein WEA) with antibody activity against 3,4-pyruvylated  
RT galactose in Klebsiella polysaccharides K30 and K33";

```

RL  Proc. Natl. Acad. Sci. U.S.A. 80:4837-4841(1983).
CC  -1- MISCELLANEOUS: THIS CHAIN WAS OBTAINED FROM A MONOCLONAL ANTIBODY
CC  AGAINST 3,4-PYRIDYLATED GALACTOSE AND ISOLATED FROM A PATIENT WITH
CC  WALDENSTROM'S MACROGLOBULINEMIA.
CC  -1- SIMILARITY: Contains 1 immunoglobulin-like domain.
DR  PIR; A02046; M3HUM.
DR  HSSP; P01772; 2FB4.
DR  GO; GO:0005576; C:extracellular; NAS.
DR  GO; GO:0003823; F:antigen binding; NAS.
DR  GO; GO:0006955; P:immune response; NAS.
DR  InterPro; IPR007110; Ig-like.
DR  InterPro; IPR003596; Ig_v.
DR  Pfam; PF00047; Ig_1.
DR  SMART; SM00406; IGV; 1.
DR  PROSITE; PS50835; IGV_LIKE; 1.
KW  Immunoglobulin V region; Pyroliidone carboxylic acid.
FT  DOMAIN 1 112
FT  NON_TER 114 114 PYROLIIDONE CARBOXYLIC ACID.
SQ  SEQUENCE 114 AA; 12256 MW; D88294FB418A07B7 CRC64;

Query Match 49.1%; Score 338; DB 1; Length 114;
Best Local Similarity 57.9%; Pred. No. 1.5e-25;
Matches 73; Conservative 15; Mismatches 22; Indels 16; Gaps 3;

QY 1 QVQLQSGGGVYVAGSGLRISCAASGISISINVMGFRAQPGKORELVASITSGGST-NY 59
DB 1 QVQLVSGGGVYVAGSGLRISCAASGISISINVMGFRAQPGKORELVASITSGGSTITY 60
QY 60 ADSLKGRFTISRDNKAKAVYLQNNMLKPEDTAVYYCAHITPAGSSNYGYG--WGHGT 116
DB 61 ADSVKGKFTISRDSKSTLYLQNNMLKPEDTAVYYCAHITPAGSSNYGYG--WGHGT 108
QY 117 KVTYSS 122
DB 109 LVTYSS 114

RESULT 7
HV3L_HUMAN STANDARD; PRT; 122 AA.
ID HV3L_HUMAN
AC P01768;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ig heavy chain V-III region CAM.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCBI_TaxID=9606;
RX MEDLINE=81013859; PubMed=6774332;
RA Lehman D.W., Putnam F.W.;
RT "Amino acid sequence of the variable region of a human mu chain:
RT location of a possible JH segment.";
RL Proc. Natl. Acad. Sci. U.S.A. 77:3239-3243(1980).
CC -1- MISCELLANEOUS: THIS MU CHAIN WAS ISOLATED FROM THE PLASMA OF A
CC PATIENT WITH MACROGLOBULINEMIA.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.
DR  PIR; A02051; M3HUM.
DR  HSSP; P01772; 2FB4.
DR  GO; GO:0005576; C:extracellular; NAS.
DR  GO; GO:0003823; F:antigen binding; NAS.
DR  GO; GO:0006955; P:immune response; NAS.
DR  InterPro; IPR007110; Ig-like.
DR  InterPro; IPR003596; Ig_v.
DR  Pfam; PF00047; Ig_1.
DR  SMART; SM00406; IGV; 1.
DR  PROSITE; PS50835; IGV_LIKE; 1.
KW  Immunoglobulin V region; Pyroliidone carboxylic acid.
FT  DOMAIN 1 112
FT  MOD_RES 1 112
FT  NON_TER 114 114 PYROLIIDONE CARBOXYLIC ACID.
SQ  SEQUENCE 114 AA; 12256 MW; D88294FB418A07B7 CRC64;

Query Match 49.1%; Score 338; DB 1; Length 114;
Best Local Similarity 57.9%; Pred. No. 1.5e-25;
Matches 73; Conservative 15; Mismatches 22; Indels 16; Gaps 3;

```

```

FT  NON_TER 122 122
SQ  SEQUENCE 122 AA; 13668 MW; A42D0F17D252F1C2 CRC64;

Query Match 49.0%; Score 337; DB 1; Length 122;
Best Local Similarity 57.7%; Pred. No. 2e-25;
Matches 71; Conservative 14; Mismatches 36; Indels 2; Gaps 2;

QY 1 QVQLQSGGGVYVAGSGLRISCAASGISISINVMGFRAQPGKORELVASITSGGSTNY 59
DB 1 QVQLVSGGGVYVAGSGLRISCAASGISISINVMGFRAQPGKORELVASITSGGSTITY 60
QY 60 ADSLKGRFTISRDNKAKAVYLQNNMLKPEDTAVYYCAHITPAGSSNYGYG--WGHGT 119
DB 61 ADSVKGKFTISRDSKSTLYLQNNMLKPEDTAVYYCAHITPAGSSNYGYG--WGHGT 116
QY 120 VSS 122
DB 120 VSS 122

RESULT 8
HV3L_HUMAN STANDARD; PRT; 119 AA.
ID HV3L_HUMAN
AC P01773;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ig heavy chain V-III region BUR.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCBI_TaxID=9606;
RX MEDLINE=79151016; PubMed=107164;
RA Putnam F.W., Liu Y.-S.V., Low T.L.R.;
RT "Primary structure of a human IgA1 immunoglobulin. IV. Streptococcal
RT IgA1 protease, digestion, Fab and Fc fragments, and the complete
RT amino acid sequence of the alpha 1 heavy chain.";
RL J. Biol. Chem. 254:2865-2874(1979).
CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.
DR  PIR; A02056; A1HUR.
DR  HSSP; P01772; 2FB4.
DR  GO; GO:0005576; C:extracellular; NAS.
DR  GO; GO:0003823; F:antigen binding; NAS.
DR  GO; GO:0006955; P:immune response; NAS.
DR  InterPro; IPR007110; Ig-like.
DR  InterPro; IPR003596; Ig_v.
DR  Pfam; PF00047; Ig_1.
DR  SMART; SM00406; IGV; 1.
DR  PROSITE; PS50835; IGV_LIKE; 1.
KW  Immunoglobulin V region; Glycoprotein; Pyroliidone carboxylic acid.
FT  DOMAIN 1 112
FT  MOD_RES 1 112
FT  DISULFID 22 96
FT  CARBOHYD 28 28 N-LINKED (GLCNAC. . .)
FT  NON_TER 119 119
SQ  SEQUENCE 119 AA; 12981 MW; 12A709A75344D024 CRC64;

Query Match 48.8%; Score 335.5; DB 1; Length 119;
Best Local Similarity 59.0%; Pred. No. 2.7e-25;
Matches 72; Conservative 14; Mismatches 31; Indels 5; Gaps 2;

```



Db 117 VS 118

## RESULT 9

ID HV3P HUMAN STANDARD; PRT; 119 AA.

AC P01777;

DT 21-JUL-1986 (Rel. 01, Created)

DT 21-JUL-1986 (Rel. 01, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Ig heavy chain V-II region TEI.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

OX NCB1\_TaxID=9606;

RN [1]

RX MEDLINE=74142702; PubMed=4522793;

RA Capra J.D., Kehoe J.M.

RT "Variable region sequences of five human immunoglobulin heavy chains

RT of the VH3 subgroup: definitive identification of four heavy chain

RT hypervariable regions."

RL Proc. Natl. Acad. Sci. U.S.A. 71:845-848 (1974).

CC -1- MISCELLANEOUS: THIS CHAIN WAS ISOLATED FROM AN IG31 MYELOMA

CC PROTEIN.

CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.

CC PIR; A02060; GIHUTE.

CC DR HSSP; P01772; 2FB4.

CC DR GO; GO:0005576; C:extracellular; NAS.

CC DR GO; GO:0003823; F:antigen binding; NAS.

CC DR GO; GO:0006955; P:immune response; NAS.

CC DR InterPro; IPR007110; IG-like.

CC DR InterPro; IPR003596; IG\_v.

CC DR Pfam; PF00047; IG\_1.

CC DR SMART; SM00406; IGV; 1.

CC DR PROSITE; PS50835; IG-LIKE; 1.

CC DR Immunoglobulin V region.

CC KW DOMAIN 1 107

CC FT NON TER 119

CC SQ SEQUENCE 119 AA; 12862 MW; 7E24DC852C7290A9 CRC64;

SQ

Query Match 48.5%; Score 334; DB 1; Length 119;

Best Local Similarity 59.2%; Pred. No. 3.8e-25;

Matches 71; Conservative 16; Mismatches 31; Indels 2; Gaps 2;

QY 1 QVQLQESGGGVQAGSRLSCAASGSISSINVMGFQAPGKRELVASITSGGS-TNY 59

DB 1 EVQLVESGGGLVQPGGSLRLSCAASGFTFTSAVYWRQAPGKGLRWGMYGSSSLTHY 60

QY 60 ADSLKGFTISRDNAKAVYLQMNLIKPEDTAVVYCNHITPAGSSNYVYGWGHGTYKT 119

DB 61 AVSVQGRFTISRDNSKNTLYQMISLEPZBETAVYVC-ARVFPAAALTFSAVMCGTLVT 119

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

CC -1- MISCELLANEOUS: THE SEQUENCES OF THE V REGIONS OF THE HEAVY CHAINS

CC OF IGM AND IG22 ISOLATED FROM A SINGLE PATIENT WITH BICLONAL

CC GAMOPATHY ARE IDENTICAL. THEIR LIGHT CHAINS ARE APPARENTLY ALSO

CC IDENTICAL.

CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.

CC PIR; A02048; H3HUTL.

CC DR HSSP; P01772; 2FB4.

CC DR GO; GO:0005576; C:extracellular; NAS.

CC DR GO; GO:0003823; F:antigen binding; NAS.

CC DR GO; GO:0006955; P:immune response; NAS.

CC DR InterPro; IPR007110; IG-like.

CC DR InterPro; IPR003596; IG\_v.

CC DR Pfam; PF00047; IG\_1.

CC DR SMART; SM00406; IGV; 1.

CC DR PROSITE; PS50835; IG-LIKE; 1.

CC DR Immunoglobulin V region.

CC KW DOMAIN 1 108

CC FT NON TER 115

CC SQ SEQUENCE 115 AA; 12356 MW; 4DCC67D179F62326 CRC64;

SQ

Query Match 48.5%; Score 333.5; DB 1; Length 115;

Best Local Similarity 59.3%; Pred. No. 4.1e-25;

Matches 73; Conservative 11; Mismatches 30; Indels 9; Gaps 3;

QY 1 QVQLQESGGGVQAGSRLSCAASGSISSINVMGFQAPGKRELVASITSGGS-TNY 59

DB 1 EVQLVESGGGLVQPGGSLRLSCAASGFTFTSAVYWRQAPGKGLRWGMYGSSSLTHY 60

QY 60 ADSLKGFTISRDNAKAVYLQMNLIKPEDTAVVYCNHITPAGSSNYVYGWGHGTYKT 119

DB 61 AVSVQGRFTISRDNSKNTLYQMISLEPZBETAVYVC-ARVFPAAALTFSAVMCGTLVT 112

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.

CC PIR; A02811; AVMS82.

CC DR HSSP; P01810; 2FBJ.

CC DR InterPro; IPR007110; IG-like.

CC DR InterPro; IPR003596; IG\_v.

CC DR Pfam; PF00047; IG\_1.

CC DR SMART; SM00406; IGV; 1.

CC DR PROSITE; PS50835; IG-LIKE; 1.

CC DR Immunoglobulin V region.

CC KW DOMAIN 1 114

CC FT DISULFID 98

CC FT NON TER 115

CC SQ SEQUENCE 115 AA; 12887 MW; 9B4517648C121GSA CRC64;

SQ

Query Match 48.2%; Score 331.5; DB 1; Length 115;

Best Local Similarity 59.3%; Pred. No. 4.1e-25;

Matches 73; Conservative 11; Mismatches 30; Indels 9; Gaps 3;

QY 1 QVQLQESGGGVQAGSRLSCAASGSISSINVMGFQAPGKRELVASITSGGS-TNY 59

DB 1 EVQLVESGGGLVQPGGSLRLSCAASGFTFTSAVYWRQAPGKGLRWGMYGSSSLTHY 60

QY 60 ADSLKGFTISRDNAKAVYLQMNLIKPEDTAVVYCNHITPAGSSNYVYGWGHGTYKT 119

DB 61 AVSVQGRFTISRDNSKNTLYQMISLEPZBETAVYVC-ARVFPAAALTFSAVMCGTLVT 112

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

Best Local Similarity 53.6%; Pred. No. 6.3e-25; Indels 13; Gaps 2;  
Matches 67; Conservative 20; Mismatches 25;

QY 1 QVQLQSGGGGLVQAGSLRLSCAASGSISSINVMGFAPGKQRELVASI---TSGGST 57  
ID HV32 HUMAN STANDARD; PRT; 116 AA.  
AC P01781;  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DE 10-OCT-2003 (Rel. 42, Last annotation update)  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP MEDLINE=75059123; PubMed=4803843;  
RA Watanabe S., Barinkol H.V., Horn J., Bertram J., Hilschmann N.;  
RT "The primary structure of a monoclonal IgM-immunoglobulin  
type), subgroup H III. Architecture of the complete IgM-molecule."  
RT Hoppe-Sejler's Z. Physiol. Chem. 354:1505-1509 (1973).  
RN [2]  
RP REVISION TO 28-33.  
RA Hilschmann N.;  
RL Submitted (JUN-1975) to the PIR data bank.  
CC -1- MISCELLANEOUS: THIS MY CHAIN WAS ISOLATED FROM A WALDENSTROM'S  
MACROGLOBULIN.  
CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.  
DR PIR; A02064; M5HUGL.  
DR HSSP; P01772; 2F84.  
DR GO; GO:0005576; C:extracellular; NAS.  
DR GO; GO:0003823; P:antigen binding; NAS.  
DR GO; GO:0006955; P:immune response; NAS.  
DR InterPro; IPR007110; IG-like.  
DR InterPro; IPR003596; IG\_V.  
DR Pfam; PF00047; IG\_1.  
DR SMART; SM00406; IGV; 1.  
DR PROSITE; PS50835; IG-LIKE; 1.  
KW Immunoglobulin V region.  
FT DOMAIN 1 112  
FT NON TER 116 116 IG-LIKE.  
SQ SEQUENCE 116 AA; 12730 MW; 2C67CA9AAAAA1282 CRC64;

Query Match 48.0%; Score 330; DB 1; Length 116;  
Best Local Similarity 61.0%; Pred. No. 8.9e-25;  
Matches 75; Conservative 10; Mismatches 30; Indels 8; Gaps 3;

QY 1 QVQLQSGGGGLVQAGSLRLSCAASGSISSINVMGFAPGKQRELVASI---TSGGST 57  
ID HV32 HUMAN STANDARD; PRT; 116 AA.  
AC P01781;  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DE 10-OCT-2003 (Rel. 42, Last annotation update)  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP MEDLINE=75059123; PubMed=4803843;  
RA Watanabe S., Barinkol H.V., Horn J., Bertram J., Hilschmann N.;  
RT "The primary structure of a monoclonal IgM-immunoglobulin  
type), subgroup H III. Architecture of the complete IgM-molecule."  
RT Hoppe-Sejler's Z. Physiol. Chem. 354:1505-1509 (1973).  
RN [2]  
RP REVISION TO 28-33.  
RA Hilschmann N.;  
RL Submitted (JUN-1975) to the PIR data bank.  
CC -1- MISCELLANEOUS: THIS MY CHAIN WAS ISOLATED FROM A WALDENSTROM'S  
MACROGLOBULIN.  
CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.  
DR PIR; A02064; M5HUGL.  
DR HSSP; P01772; 2F84.  
DR GO; GO:0005576; C:extracellular; NAS.  
DR GO; GO:0003823; P:antigen binding; NAS.  
DR GO; GO:0006955; P:immune response; NAS.  
DR InterPro; IPR007110; IG-like.  
DR InterPro; IPR003596; IG\_V.  
DR Pfam; PF00047; IG\_1.  
DR SMART; SM00406; IGV; 1.  
DR PROSITE; PS50835; IG-LIKE; 1.  
KW Immunoglobulin V region.  
FT DOMAIN 1 112  
FT NON TER 116 116 IG-LIKE.  
SQ SEQUENCE 116 AA; 12730 MW; 2C67CA9AAAAA1282 CRC64;

## RESULT 13

QY 1 QVQLQSGGGGLVQAGSLRLSCAASGSISSINVMGFAPGKQRELVASI---TSGGST 57  
ID HV32 MOUSE STANDARD; PRT; 115 AA.  
AC P01801;  
DT 21-JUL-1986 (Rel. 01, Created)  
DE 10-OCT-2003 (Rel. 42, Last sequence update)  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP MEDLINE=82099361; PubMed=6798111;  
RA Johnson N., Stankard J., Paul L., Hood L.;  
RT "The complete V domain amino acid sequences of two myeloma inulin-  
binding proteins."  
RT J. Immunol. 128:302-307 (1982).  
CC -1- MISCELLANEOUS: THIS CHAIN WAS ISOLATED FROM A MYELOMA PROTEIN THAT  
BINDS INULIN.  
CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.  
DR PIR; C92811; AVMS06.  
DR HSSP; P01789; IMCP.  
DR InterPro; IPR007110; IG-like.  
DR InterPro; IPR003596; IG\_V.  
DR Pfam; PF00047; IG\_1.  
DR SMART; SM00406; IGV; 1.  
DR PROSITE; PS50835; IG-LIKE; 1.  
KW Immunoglobulin V region.  
FT DOMAIN 1 114  
FT DISULFID 22 98  
FT NOW TER 115 115 BY SIMILARITY.  
SQ SEQUENCE 115 AA; 12810 MW; B67AD6638A12125F CRC64;

Query Match 47.3%; Score 329.5; DB 1; Length 115;  
Best Local Similarity 53.6%; Pred. No. 9.8e-25;  
Matches 67; Conservative 20; Mismatches 25; Indels 13; Gaps 2;

QY 1 QVQLQSGGGGLVQAGSLRLSCAASGSISSINVMGFAPGKQRELVASI---TSGGST 57  
ID HV32 HUMAN STANDARD; PRT; 116 AA.  
AC P01781;  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DE 10-OCT-2003 (Rel. 42, Last annotation update)  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP MEDLINE=75059123; PubMed=4803843;  
RA Watanabe S., Barinkol H.V., Horn J., Bertram J., Hilschmann N.;  
RT "The primary structure of a monoclonal IgM-immunoglobulin  
type), subgroup H III. Architecture of the complete IgM-molecule."  
RT Hoppe-Sejler's Z. Physiol. Chem. 354:1505-1509 (1973).  
RN [2]  
RP REVISION TO 28-33.  
RA Hilschmann N.;  
RL Submitted (JUN-1975) to the PIR data bank.  
CC -1- MISCELLANEOUS: THIS MY CHAIN WAS ISOLATED FROM A WALDENSTROM'S  
MACROGLOBULIN.  
CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.  
DR PIR; A02064; M5HUGL.  
DR HSSP; P01772; 2F84.  
DR GO; GO:0005576; C:extracellular; NAS.  
DR GO; GO:0003823; P:antigen binding; NAS.  
DR GO; GO:0006955; P:immune response; NAS.  
DR InterPro; IPR007110; IG-like.  
DR InterPro; IPR003596; IG\_V.  
DR Pfam; PF00047; IG\_1.  
DR SMART; SM00406; IGV; 1.  
DR PROSITE; PS50835; IG-LIKE; 1.  
KW Immunoglobulin V region.  
FT DOMAIN 1 112  
FT NON TER 116 116 IG-LIKE.  
SQ SEQUENCE 116 AA; 12730 MW; 2C67CA9AAAAA1282 CRC64;

## RESULT 14

QY 1 QVQLQSGGGGLVQAGSLRLSCAASGSISSINVMGFAPGKQRELVASI---TSGGST 57  
ID HV02 CANPA STANDARD; PRT; 117 AA.  
AC P01785;  
DT 21-JUL-1986 (Rel. 01, Created)  
DE 10-OCT-2003 (Rel. 42, Last sequence update)  
OS Canis familiaris (Dog).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.  
OX NCBI\_TaxID=9615;  
RN [1]  
RP MEDLINE=77242268; PubMed=407924;  
RA Messerman R.L., Capra J.D.;  
RT "Primary structure of the variable regions of two canine  
immunoglobulin heavy chains."  
RT Biochemistry 16:3160-3168 (1977).  
RN [2]  
RP REVISION TO 28-33.  
RA Hilschmann N.;  
RL Submitted (JUN-1975) to the PIR data bank.  
CC -1- MISCELLANEOUS: THIS MY CHAIN WAS ISOLATED FROM A WALDENSTROM'S  
MACROGLOBULIN.  
CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.  
DR PIR; A02064; M5HUGL.  
DR HSSP; P01772; 2F84.  
DR GO; GO:0005576; C:extracellular; NAS.  
DR GO; GO:0003823; P:antigen binding; NAS.  
DR GO; GO:0006955; P:immune response; NAS.  
DR InterPro; IPR007110; IG-like.  
DR InterPro; IPR003596; IG\_V.  
DR Pfam; PF00047; IG\_1.  
DR SMART; SM00406; IGV; 1.  
DR PROSITE; PS50835; IG-LIKE; 1.  
KW Immunoglobulin V region.  
FT DOMAIN 1 114  
FT DISULFID 22 98  
FT NOW TER 115 115 BY SIMILARITY.  
SQ SEQUENCE 115 AA; 12810 MW; B67AD6638A12125F CRC64;

RN	[2]	RP	SEQUENCE OF 113-117.	
RX		MEDLINE=80077682; PubMed=117299;		
RA		McCumber L.J., Capra J.D.,		
RT		"The complete amino-acid sequence of a canine mu chain."		
RL		Mol. Immunol. 16:565-570(1979).		
CC		-1- MISCELLANEOUS: THIS MU CHAIN WAS ISOLATED FROM A MYELOMA PROTEIN.		
DR		-1- SIMILARITY: Contains 1 immunoglobulin-like domain.		
DR		PIR; A90403; MEDGMO.		
DR		HSSP; P01772; 2F84.		
DR		InterPro; IPR007110; IG-1like.		
DR		InterPro; IPR003596; IG_V.		
DR		Pfam; PF00047; IG_1.		
DR		SMART; SM00406; IGV_1.		
KW		PROSITE; PS50835; IG_LIKE; 1.		
KW		Immunoglobulin V region.		
FT	DOMAIN	1		
FT	NON TER	117		
FT		117		
SO	SEQUENCE	117 AA; 12703 MW; FCE07309E0A64B35 CRC64;		

This Page Blank (uspio)



Db 1 EVOLVSSGGGLVQPGSLRLSCAASGFTVSSNVMWVRQAPGKGLSESVVYSGGSYYA 60  
 QY 61 DSLKGRFTISRDNKNAAYVLQNNLKPEPTAVYYCNAAHTPAGSSNYYGVGHGTRV 120  
 Db 61 DSVKGRFTISRDNKNTLYLQNNLSLRADPTAVYCC---AADRFGFLFDYWGQGLTAV 116  
 QY 121 SS 122  
 Db 117 SS 118

## RESULT 2

Q8N5K4 PRELIMINARY; PRT; 499 AA.

AC 08N5K4;  
 ID 08N5K4;  
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Hypothetical protein.  
 OS Homo sapiens (human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Blood;  
 RA Strausberg R.;  
 RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: BC032249; AAH32249.1; -.  
 DR InterPro: IPR003599; IG.  
 DR InterPro: IPR007110; IG-like.  
 DR InterPro: IPR003597; IG\_c1.  
 DR InterPro: IPR003006; IG\_MHC.  
 DR InterPro: IPR003596; IG\_v.  
 DR Pfam: PF00047; IG\_4.  
 DR SMART: SM00409; IG\_4.  
 DR SMART: SM00407; IGc1.2.  
 DR SMART: SM00406; IGv.1.  
 DR PROSITE: PS00835; IG\_LIKE; 4.  
 DR PROSITE: PS00290; IG\_MHC; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 499 AA; 53376 MW; 93A5C99582054F32 CRC64;

Query Match 55.1%; Score 379; DB 4; Length 499;

Best Local Similarity 60.6%; Pred. No. 2.2e-27;  
 Matches 83; Conservative 13; Mismatches 35; Indels 6; Gaps 3;

QY 1 QVQLQESGGGLVQAGSLRLSCAASGSISSINVMGFRQAPGKORELVASIT-SGGSITNY 59  
 Db 20 EVQLVSSGGGLVQPGSLRLSCATSGFTDDSGASVVRQAPGKGLFWVSITIMNGGSITNY 79  
 QY 60 ADSLKGRFTISRDNKNAAYVLQNNLKPEPTAVYYCNAAHTPAGSSNYYGV---WGHG 115  
 Db 80 ADSVKGRTISRDNKNSLYLQNNLSLRADPTAVYCCARDPFKYCGSGSGCLGYMDVMWKG 139  
 QY 116 TKYTVSS-EPKTPKPPQ 131  
 Db 140 TTYTVSSASPTSKVPF 156

## RESULT 3

Q96BB9 PRELIMINARY; PRT; 597 AA.

AC 096BB9;  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Hypothetical protein.  
 OS Homo sapiens (human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 NCBI\_TaxID=9606;  
 RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=B-cell;  
 RA Strausberg R.;  
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: BC015760; AAH15760.1; -.  
 DR InterPro: IPR007110; IG-like.  
 DR InterPro: IPR003006; IG\_MHC.  
 DR InterPro: IPR003596; IG\_v.  
 DR Pfam: PF00047; IG\_5.  
 DR SMART: SM00406; IGv.1.  
 DR PROSITE: PS00835; IG\_LIKE; 5.  
 DR PROSITE: PS00290; IG\_MHC; 3.  
 KW Hypothetical protein.  
 SQ SEQUENCE 597 AA; 65039 MW; 4FCA3ADBEC263D9 CRC64;

Query Match 54.9%; Score 377.5; DB 4; Length 597;

Best Local Similarity 65.4%; Pred. No. 3.8e-27;  
 Matches 83; Conservative 11; Mismatches 26; Indels 7; Gaps 4;

QY 1 QVQLQESGGGLVQAGSLRLSCAASGSISSINVMGFRQAPGKORELVASIT-SGGSITNY 59  
 Db 20 EVQLVSSGGGLVQPGSLRLSCAASGFSFSYAMWVRQAPGKGLFWGSAISGGSITNY 79  
 QY 60 ADSLKGRFTISRDNKNAAYVLQNNLKPEPTAVYYCNAAHTPAG--SSNYY-YGVWGHG 115  
 Db 80 ADSVKGRTISRDNKNTLYLQNNLSLRADPTAVYCCAR--DRGYSAGNTREDYWGQ 137  
 QY 116 TKYTVSS 122  
 Db 138 TLVTVSS 144

## RESULT 4

Q8WUJ8 PRELIMINARY; PRT; 573 AA.

AC 08WUJ8;  
 ID 08WUJ8;  
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Hypothetical protein.  
 OS Homo sapiens (human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=tonsil;  
 RA Strausberg R.;  
 RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: BC021276; AAH21276.1; -.  
 DR PIR: S21205; S21205.  
 DR InterPro: IPR007110; IG-like.  
 DR InterPro: IPR003006; IG\_MHC.  
 DR InterPro: IPR003596; IG\_v.  
 DR Pfam: PF00047; IG\_4.  
 DR SMART: SM00406; IGv.1.  
 DR PROSITE: PS00835; IG\_LIKE; 4.  
 DR PROSITE: PS00290; IG\_MHC; 2.  
 KW Hypothetical protein.  
 SQ SEQUENCE 573 AA; 62967 MW; FD072344033AC530 CRC64;

Query Match 54.6%; Score 375.5; DB 4; Length 573;

Best Local Similarity 60.9%; Pred. No. 5.6e-27;  
 Matches 84; Conservative 12; Mismatches 31; Indels 11; Gaps 5;

QY 1 QVQLQESGGGLVQAGSLRLSCAASGSISSINVMGFRQAPGKORELVASIT-SGGSITNY 59  
 Db 20 EVQLVSSGGGLVQPGSLRLSCAASGFTDDYAMWVRQAPGKGLFWGSAISGGSITNY 79  
 QY 60 ADSLKGRFTISRDNKNAAYVLQNNLKPEPTAVYYCNAAHTPAGSSNYY---YGVWGHG 114  
 Db 80 ADSVKGRTISRDNKNSLYLQNNLSLRADPTAVYCCAR---GSGSYGYGYMDVMWKG 135

```

OY 115 GTRVTSSEP-KTTPKPOP 131
DB 136 GTTVTSAPTKADVPF 153

RESULT 5
O96K68 PRELIMINARY; PRT; 494 AA.
ID Q96K68:
AC Q96K68:
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Hypothetical protein FLJ14473.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC Tissue=Mammary gland;
RA Isega T., Oca T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
RA Takahashi M., Hosoi T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA Yamanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahara K., Masuko Y.,
RA Niimoriya K., Iwayanagi T.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK027379; BAB55072.1; -.
DR PIR; S21205; S21205.
DR InterPro; IPR007110; IG_1like.
DR InterPro; IPR003006; IG_MHC.
DR InterPro; IPR003596; IG_V.
DR Pfam; PF00047; Ig; 4.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS00835; IG_Like; 4.
DR PROSITE; PS00290; IG_MHC; 1.
KW Hypothetical protein.
SQ SEQUENCE 494 AA; 53088 MW; 9A1D7AEB5AE4C0E CRC64;

Query Match 52.8%; Score 363.5; DB 4; Length 494;
Best Local Similarity 60.9%; Pred. No. 6.3e-26;
Matches 81; Conservative 14; Mismatches 35; Indels 3; Gaps 3;

OY 1 QVQLQESGGGLVQAGSLRLSCAASGSISSINVMGFWFRQAPGKORELVAITSGGS-TNY 59
DB 20 EVQLVESGGGLVQPGGSLRLSCAASGLSPSTYAMMWVQADPGKLEWVSSISSRSDIYY 79
OY 60 ADSLKGRFTISRDNKNAVYLOMNNLKPEDTAVYYCNANHPPASSNNVYGYWGHSTKVT 119
DB 80 RDSVAKGRFTISRDNKNSLYLOMNSLRVDDTAVYYIC-ARDSCNGAICYGSPWQGITLV 138
OY 120 VSS-EPKTPKPOP 131
DB 139 VSSASPTSPKVPF 151

RESULT 6
O920E7 PRELIMINARY; PRT; 119 AA.
ID Q920E7:
AC Q920E7:
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Pterin-mimicking anti-idiotope heavy chain variable region
DE (Fragment).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RA Alkin J.D., Iape A., Jennings I.G., Horaltis O., Cotton R.G.H.;
RT "Definition of the idiotope of Pterin-Mimicking Antibodies Expressed
RT in Mammalian Cells.";
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF307937; AAL09421.1; -.
DR PIR; C25913; C25913.
DR InterPro; IPR007110; IG_1like.
DR InterPro; IPR003596; IG_V.
DR Pfam; PF00047; Ig; 1.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS00835; IG_Like; 1.
DR PROSITE; PS00290; IG_MHC; 1.
FT NON_TER 1
FT NON_TER 119
SQ SEQUENCE 119 AA; 13025 MW; F6B90404381CA7C CRC64;

Query Match 52.7%; Score 362.5; DB 11; Length 119;
Best Local Similarity 61.8%; Pred. No. 1.3e-26;
Matches 76; Conservative 14; Mismatches 28; Indels 5; Gaps 2;

OY 1 QVQLQESGGGLVQAGSLRLSCAASGSISSINVMGFWFRQAPGKORELVAITSGGS-TNY 59
DB 1 EVQLVESGGGLVQPGGSLRLSCAASGFTPSYGMSTVPRQTPKRLRWVAITISGGSYTY 60
OY 60 ADSLKGRFTISRDNKNAVYLOMNNLKPEDTAVYYCNANHPPASSNNVYGYWGHSTKVT 119
DB 61 PDSVAKGRFTISRDNKNSLYLOMNSLRVDDTAVYYICARH---GDYDGFAYWGQITLV 116
OY 120 VSS 122
DB 117 VSA 119

RESULT 7
O9UL93 PRELIMINARY; PRT; 116 AA.
ID Q9UL93:
AC Q9UL93:
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Myosin-reactive immunoglobulin heavy chain variable region
DE (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=8277139; PubMed=9614934;
RA Wu X., Liu B., Van der Werf P.L., Kallis N.N., Berney S.M.,
RA Young D.C.;
RT "Myosin-reactive autoantibodies in rheumatic carditis and normal
RT fetus.";
RL Clin. Immunol. Immunopathol. 87:184-192 (1998).
DR EMBL; AF035021; AAD56257.1; -.
DR PIR; PLO120; PLO120.
DR HSSP; P01772; 2F84.
DR InterPro; IPR007110; IG_1like.
DR InterPro; IPR003596; IG_V.
DR Pfam; PF00047; Ig; 1.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS00835; IG_Like; 1.
FT NON_TER 1
FT NON_TER 116
SQ SEQUENCE 116 AA; 12434 MW; ODA0348154DD6061 CRC64;

Query Match 52.8%; Score 361.5; DB 4; Length 116;
Best Local Similarity 63.9%; Pred. No. 1.6e-26;
Matches 78; Conservative 8; Mismatches 29; Indels 7; Gaps 2;

OY 2 VQVLESQGLVQAGSLRLSCAASGSISSINVMGFWFRQAPGKORELVAITSGGS-TNY 60
DB 1 VQLVESGGGLVQPGGSLRLSCAASGFTPSYGMSTVPRQTPKRLRWVAITISGGSYTY 60

```

QY 61 DSLKGRFTISRDNAAVYLOMNNLKPEDTAVYYCNAHITPAGSSNYYGYWGHGHTKTV 120  
 DB 61 DSLKGRFTISRDNAAVYLOMNNLKPEDTAVYYCNAHITPAGSSNYYGYWGHGHTKTV 114  
 QY 121 SS 122  
 DB 115 SS 116

## RESULT 8

Q99KA4 PRELIMINARY; PRT; 487 AA.  
 ID Q99KA4;  
 AC Q99KA4;  
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Hypochemical protein.  
 OS Mus musculus (mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NC NCB1\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RA Strausberg R.  
 RL Submitted (MAR-2001) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; BC004786; AA04786.1; -  
 DR HSSP; P01810; 2FB0.  
 DR InterPro; IPR007110; IG-like.  
 DR InterPro; IPR003006; IG\_MHC.  
 DR InterPro; IPR003596; IG\_v.  
 DR Pfam; PF00047; IG\_4.  
 DR SMART; SM00406; IGv\_1.  
 DR PROSITE; PS50835; IG\_LIKE; 4.  
 DR PROSITE; PS00290; IG\_MHC; 2.  
 KW Hypochemical protein.  
 SQ SEQUENCE 487 AA; 52554 MW; 7DC8E96DB333077B CRC64;

Query Match 52.5%; Score 361; DB 11; Length 487;  
 Best Local Similarity 60.2%; Pred. No. 1.1e-25;  
 Matches 77; Conservative 15; Mismatches 28; Indels 8; Gaps 3;

QY 1 QVQLQESGGGLVQAGSLRLSCAASGSSISINVMGFQAPGKQRELVASITSGS-TNY 59  
 DB 20 EVQLVESGGGLVPGGSLRLSCAASGFTPSVAMSVVRGTPKRLRWATIDGSSYYTY 79  
 QY 60 ADSLKGFTISRDNAAVYLOMNNLKPEDTAVYYCNAHITPAGSSNYYGYWGHG 115  
 DB 80 PDVVKRFTISRDNAAVYLOMNNLKPEDTAVYYCNAHITPAGSSNYYGYWGHG 136  
 QY 116 TKVTVSSSE 123  
 DB 137 TITVSSSE 144

## RESULT 9

Q9UL91 PRELIMINARY; PRT; 118 AA.  
 ID Q9UL91;  
 AC Q9UL91;  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Myosin-reactive immunoglobulin heavy chain variable region  
 DE (Fragment).  
 OS Homo sapiens (human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 NC NCB1\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RX MEDLINE=98277139; PubMed=9614934;  
 RA Wu X., Liu B., Van der Werwe P.L., Kalis N.N., Berney S.M.,  
 RA Young D.C.;  
 RT "Myosin-reactive autoantibodies in rheumatic carditis and normal

RT "fetus";  
 RL Clin. Immunol. Immunopathol. 87:184-192 (1998).  
 DR EMBL; AF035023; AAD56259.1; -  
 DR PIR; S21205; S21205.  
 DR HSSP; P01772; 2FB4.  
 DR InterPro; IPR007110; IG-like.  
 DR InterPro; IPR003596; IG\_v.  
 DR Pfam; PF00047; IG\_1.  
 DR SMART; SM00406; IGv\_1.  
 DR PROSITE; PS50835; IG\_LIKE; 1.  
 FT NON\_TER  
 FT NON\_TER  
 SQ SEQUENCE 118 AA; 12843 MW; D0633949F2AC19D CRC64;

Query Match 52.0%; Score 358; DB 4; Length 118;  
 Best Local Similarity 63.3%; Pred. No. 3.5e-26;  
 Matches 78; Conservative 10; Mismatches 28; Indels 6; Gaps 2;

QY 1 QVQLQESGGGLVQAGSLRLSCAASGSSISINVMGFQAPGKQRELVASITSGS-TNY 59  
 DB 1 EVQLVESGGGLVPGGSLRLSCAASGFTPSVAMSVVRGTPKRLRWATIDGSSYYTY 60  
 QY 60 ADSLKGFTISRDNAAVYLOMNNLKPEDTAVYYCNAHITPAGSSNYYGYWGHGHTKTV 119  
 DB 61 ADSLKGFTISRDNAAVYLOMNNLKPEDTAVYYCNAHITPAGSSNYYGYWGHGHTKTV 115  
 QY 120 VS 121  
 DB 116 VS 117

## RESULT 10

Q9UL90 PRELIMINARY; PRT; 113 AA.  
 ID Q9UL90;  
 AC Q9UL90;  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Myosin-reactive immunoglobulin heavy chain variable region  
 DE (Fragment).  
 OS Homo sapiens (human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 NC NCB1\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RX MEDLINE=98277139; PubMed=9614934;  
 RA Wu X., Liu B., Van der Werwe P.L., Kalis N.N., Berney S.M.,  
 RA Young D.C.;  
 RT "Myosin-reactive autoantibodies in rheumatic carditis and normal  
 RT fetus";  
 RL Clin. Immunol. Immunopathol. 87:184-192 (1998).  
 DR EMBL; AF035024; AAD56260.1; -  
 DR HSSP; P01772; 2FB4.  
 DR InterPro; IPR007110; IG-like.  
 DR InterPro; IPR003596; IG\_v.  
 DR Pfam; PF00047; IG\_1.  
 DR SMART; SM00406; IGv\_1.  
 DR PROSITE; PS50835; IG\_LIKE; 1.  
 FT NON\_TER  
 FT NON\_TER  
 SQ SEQUENCE 113 AA; 12437 MW; ED57FDD19086D07F CRC64;

Query Match 51.5%; Score 354.5; DB 4; Length 113;  
 Best Local Similarity 62.6%; Pred. No. 7.1e-26;  
 Matches 77; Conservative 8; Mismatches 27; Indels 11; Gaps 2;

QY 1 QVQLQESGGGLVQAGSLRLSCAASGSSISINVMGFQAPGKQRELVASITSGS-TNY 59  
 DB 1 EVQLVESGGGLVPGGSLRLSCAASGFTPSVAMSVVRGTPKRLRWATIDGSSYYTY 60  
 QY 60 ADSLKGFTISRDNAAVYLOMNNLKPEDTAVYYCNAHITPAGSSNYYGYWGHGHTKTV 119  
 DB 61 ADSLKGFTISRDNAAVYLOMNNLKPEDTAVYYCNAHITPAGSSNYYGYWGHGHTKTV 115



```

Db      61 ADVKGRFTISRDNKSNKTLVLOMNSLRADPTAVVYCANL-----NYWGQGLTV 110
Qy      120 VSS 122
      |||
Db      111 VSS 113

RESULT 11
Q9UL88      PRELIMINARY;      PRT;      131 AA.
AC Q9UL88;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Myosin-reactive immunoglobulin heavy chain variable region
DE (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98277139; PubMed=964934;
RA Wu X., Liu B., Van der Merwe P.L., Kalis N.N., Berney S.M.,
RA Young D.C.;
RT "Myosin-reactive autoantibodies in rheumatic carditis and normal
RT fevers."
RL Clin. Immunol. Immunopathol. 87:184-192(1998).
DR EMBL; AF035026; AAD56262.1; -.
DR PIR; S21205; S21205.
DR HSSP; P01810; 2F84.
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003596; IG_v.
DR Pfam; PF00047; IG_1.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS50835; IG_LIKE; 1.
FT NON TER
FT NON TER
SQ SEQUENCE 131 AA; 14142 MW; 96E7D668E375D8A0 CRC64;

Query Match
Best Local Similarity 51.5%; Score 354.5; DB 4; Length 131;
Matches 77; Conservative 11; Mismatches 34; Indels 9; Gaps 2;

Qy      1 QVQLQESGGGLVQAGSLRLSCAASGSISSINVMGFQAPGKRELVASI---TSGGST 57
      |||
Db      1 EVQLVESGGGLVQPGGSLRLSCAASGFTFSKAMGWVQAPGKGLHWGRISKTDGDTT 60
      |||
Qy      58 NYADSLKGRFTISRDNKSNKTLVLOMNSLRADPTAVVYCANLIT-----PAGSSNYVYGY 111
      |||
Db      61 DYAAPVKGSLTISRDNKSNKTLVLOMNSLRADPTAVVYCTTGITMITIVITTSKRTSPFY 120
      |||
Qy      112 WGGTKVTVSS 122
      |||
Db      121 WGGTKVTVSS 131

RESULT 12
Q9HCC1      PRELIMINARY;      PRT;      112 AA.
AC Q9HCC1;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Single chain Fv (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX Kikuchi M., Takeda C., Tsujimoto Y., Asada S., Nagata K.;
RT "An antibody fragment2a3 specific for native lysozyme: Isolation from a

```

```

RT      human synthetic phage display library and characterization."
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB049915; BAB16829.1; -.
DR HSSP; P01772; 2F84.
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003596; IG_v.
DR Pfam; PF00047; IG_1.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS50835; IG_LIKE; 1.
FT NON TER
FT NON TER
SQ SEQUENCE 112 AA; 12243 MW; 24FLA45EC3B84788 CRC64;

Query Match
Best Local Similarity 51.5%; Score 354; DB 4; Length 112;
Matches 73; Conservative 11; Mismatches 27; Indels 8; Gaps 2;

Qy      1 QVQLQESGGGLVQAGSLRLSCAASGSISSINVMGFQAPGKRELVASIT-SGGSTNY 59
      |||
Db      1 EVQLVESGGGLVQPGGSLRLSCAASGFTFDYKMSWVRQAPGKGLHWVSGINNGGSTGY 60
      |||
Qy      60 ADSLKGRFTISRDNKSNKTLVLOMNSLRADPTAVVYCANLITPAGSSNYVYGYGAGTKY 118
      |||
Db      61 ADVKGRFTISRDNKSNKTLVLOMNSLRADPTAVVYCANLITPAGSSNYVYGYGAGTKY 112
      |||

RESULT 13
Q91XEL      PRELIMINARY;      PRT;      480 AA.
AC Q91XEL;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Hypothetical protein (Fragment).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX Strauberg R.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC010798; AAH10798.1; -.
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003066; IG_MHC.
DR InterPro; IPR003596; IG_v.
DR Pfam; PF00047; IG_4.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS50835; IG_LIKE; 4.
DR PROSITE; PS00290; IG_MHC; 2.
KW Hypothetical protein.
FT NON TER
FT NON TER
SQ SEQUENCE 480 AA; 51936 MW; 20B9234E9F2B41ED CRC64;

Query Match
Best Local Similarity 50.9%; Score 350.5; DB 11; Length 480;
Matches 75; Conservative 13; Mismatches 29; Indels 7; Gaps 2;

Qy      2 VQLQESGGGLVQAGSLRLSCAASGSISSINVMGFQAPGKRELVASIT-SGGSTNYA 60
      |||
Db      20 VKLVESGGGLVQPGGSLRLSCAASGFTFSYVSWVRQGTPEKLEWVATISNGGYATHP 79
      |||
Qy      61 DSLKGRFTISRDNKSNKTLVLOMNSLRADPTAVVYCANLITPAGSSNYVYGYWGHGRTVY 120
      |||
Db      80 DSMKGRFTISRDNKSNKTLVLOMNSLRADPTAVVYCTTGITMITIVITTSKRTSPFY 133
      |||
Qy      121 SSEP 124
      |||
Db      134 SSEP 137

RESULT 14

```

```

09UL71
ID Q9UL71 PRELIMINARY; PRT; 121 AA.
AC Q9UL71;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Myosin-reactive immunoglobulin heavy chain variable region
DE (Fragment).
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98277139; PubMed=9614934;
RA Wu X., Liu B., Van der Merwe P.W., Kalis N.N., Berney S.W.,
RA Young D.C.;
RT "Myosin-reactive autoantibodies in rheumatic carditis and normal
RT fetus.";
RL Clin. Immunol. Immunopathol. 87:184-192(1998).
DR EMBL; AR035043; AAD56279.1; -.
DR HSP; P01772; ZPB4.
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003596; IG_v.
DR Pfam; PF00047; IG; 1.
DR SMART; SM00406; IG; 1.
DR PROSITE; PS50835; IG_LIKE; 1.
FT NON_TER 1
FT NON_TER 121
SQ SEQUENCE 121 AA; 13154 MW; 2F045CCFASD50736 CRC64;

Query Match 50.8%; Score 349.5; DB 4; Length 121;
Best Local Similarity 60.3%; Pred. No. 2.3e-25;
Matches 76; Conservative 12; Mismatches 29; Indels 9; Gaps 3;

QY 1 QVQLQESGGGLVQAGSLRLSCAASGSISSINVMGFRQAPGKQREIVASIT--SGGSTNY 59
Db 1 EVQLVSGGGGVQPGSGMLKSLCAASGFTFSNVMNVRQSPKGLAVARILRSNNVAT 60

QY 60 ADSLKGFTISRDNKNAVYLQNNLKPEDTAVYICNAHITPASSSNVYGVYGNHGHT 116
Db 61 ADSVKGFTISRDNKSNLYLQNNSLRADTALTYC-----AKGKVTITYDRFDIWGQT 115

QY 117 KTVSS 122
Db 116 MVTSS 121

RESULT 15
Q80Z17 PRELIMINARY; PRT; 487 AA.
AC Q80Z17;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Mus musculus (mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX Strauberg R.;
RA Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
RL EMBL; BC049143; AA049143.1; -.
DR InterPro; IPR003599; IG.
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003597; IG_c1.
DR InterPro; IPR003006; IG_MHC.
DR InterPro; IPR003596; IG_v.
DR Pfam; PF00047; IG; 3.
DR SMART; SM00409; IG; 3.

```

```

DR SMART; SM00407; IGc1; 3.
DR SMART; SM00406; IG; 1.
DR PROSITE; PS50835; IG_LIKE; 4.
DR PROSITE; PS00290; IG_MHC; 2.
DR Hypothetical protein.
SQ SEQUENCE 487 AA; 53019 MW; 31F2C893900A4D80 CRC64;

Query Match 50.7%; Score 348.5; DB 11; Length 487;
Best Local Similarity 55.9%; Pred. No. 1.6e-24;
Matches 71; Conservative 21; Mismatches 30; Indels 5; Gaps 2;

QY 1 QVQLQESGGGLVQAGSLRLSCAASGSISSINVMGFRQAPGKQREIVASIT--TSGGST 57
Db 20 EVKLKFNAGGGLVQPGSGMLKSLCAASGFTFSNVMNVRQSPKGLAVARILRSNNVAT 79

QY 58 NYADSLKGRFTISRDNKNAVYLQNNLKPEDTAVYICNAHITPASSSNVYGVYGNHGHT 117
Db 80 HYAESVKGFTISRDNKSNLYLQNNSLRADTALTYCTRR--GYDPMVYFDVWGAGITT 137

QY 118 VTVSSEP 124
Db 138 VTVSSEP 144

```

Search completed: October 5, 2004, 08:13:46  
 Job time : 69.341 secs

XX  
PS Example 2; Page 9; 37pp; English

XX The patent discloses antibodies or their fragments comprising a heavy  
CC chain variable domain (VH) derived from an immunoglobulin naturally  
CC devoid of light chains specific for inhibiting human dietary enzymes. The  
CC antibodies of the invention are useful for the preparation of medicaments  
CC or food for inhibiting the activity of one or more human dietary enzymes  
CC especially human pancreatic lipase (HPL) or human gastric lipase (HGL)  
CC which are useful for the cosmetic control of body weight of human beings.  
CC The present peptide sequence is HPL inhibiting VH fragment, HPL #15 from  
CC llama (camelid) species  
XX  
SQ Sequence 131 AA:  
Query Match 100.0%; Score 688; DB 4; Length 131;  
Best Local Similarity 100.0%; Pred. No. 1.6e-50;  
Matches 131; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 QVQLQESGGGLVQAGSLRLSCAASGSISSINVMGFQAPGKQRLVASITSGSTNYA 60  
DB 1 QVQLQESGGGLVQAGSLRLSCAASGSISSINVMGFQAPGKQRLVASITSGSTNYA 60  
OY 61 DSLKGRFTISRDNKAAVYIQMNNLKPEDTAVYYCAHITTPAGSSNYVYGYMGHGTQTV 120  
DB 61 DSLKGRFTISRDNKAAVYIQMNNLKPEDTAVYYCAHITTPAGSSNYVYGYMGHGTQTV 120  
OY 121 SSEPKTPKPOP 131  
DB 121 SSEPKTPKPOP 131  
RESULT 2  
ID AAE10552 standard; peptide; 130 AA.  
XX AAE10552;  
AC 10-DEC-2001 (first entry)  
XX HPL inhibiting VH fragment, HPL #12 from llama species.  
XX  
KW llama antibody; camelid; anorectic; heavy chain variable domain; VH;  
KW human dietary enzyme inhibitor; medicament; human pancreatic lipase; HPL;  
KW food; human gastric lipase; HGL; cosmetic control; body weight.  
XX  
OS Lama sp.  
XX  
FH Key Location/Qualifiers  
FT Region /label= CDR1  
FT /note= "Complementarity determining region 1"  
FT /label= CDR2  
FT /note= "Complementarity determining region 2"  
FT /label= CDR3  
FT /note= "Complementarity determining region 3"  
XX  
XX EP1134231-A1.  
XX  
XX 19-SEP-2001.  
XX  
XX 20-FEB-2001; 2001EP-00200703.  
XX  
XX 14-MAR-2000; 2000EP-00200930.  
XX  
XX (UNIL ) UNILEVER NV.  
XX (UNIL ) UNILEVER PLC.  
XX  
XX Bezemer S, Van De Burg M, De Haard JWM, Tareilus E;  
XX WPI; 2001-572718/65.  
XX  
XX New antibody or its fragments for inhibiting human dietary enzymes,  
PT

PT useful for cosmetic control of body weight of human beings, comprises  
PT heavy chain variable domain derived from immunoglobulin naturally devoid  
PT of light chains.  
XX  
XX Example 2; Page 9; 37pp; English.  
XX  
CC The patent discloses antibodies or their fragments comprising a heavy  
CC chain variable domain (VH) derived from an immunoglobulin naturally  
CC devoid of light chains specific for inhibiting human dietary enzymes. The  
CC antibodies of the invention are useful for the preparation of medicaments  
CC or food for inhibiting the activity of one or more human dietary enzymes  
CC especially human pancreatic lipase (HPL) or human gastric lipase (HGL)  
CC which are useful for the cosmetic control of body weight of human beings.  
CC The present peptide sequence is HPL inhibiting VH fragment, HPL #12 from  
CC llama (camelid) species  
XX  
SQ Sequence 130 AA:  
Query Match 77.4%; Score 532.5; DB 4; Length 130;  
Best Local Similarity 77.9%; Pred. No. 2.2e-37;  
Matches 102; Conservative 11; Mismatches 17; Indels 1; Gaps 1;  
OY 1 QVQLQESGGGLVQAGSLRLSCAASGSISSINVMGFQAPGKQRLVASITSGSTNYA 60  
DB 1 QVQLQESGGGLVQAGSLRLSCAASGSISSINVMGFQAPGKQRLVASITSGSTNYA 60  
OY 61 DSLKGRFTISRDNKAAVYIQMNNLKPEDTAVYYCAHITTPAGSSNYVYGYMGHGTQTV 120  
DB 61 DSLKGRFTISRDNKAAVYIQMNNLKPEDTAVYYCAHITTPAGSSNYVYGYMGHGTQTV 120  
OY 121 SSEPKTPKPOP 131  
DB 121 SSEPKTPKPOP 130  
RESULT 3  
ID AAE10553 standard; peptide; 129 AA.  
XX AAE10553;  
AC 10-DEC-2001 (first entry)  
XX HPL inhibiting VH fragment, HPL #13 from llama species.  
XX  
KW llama antibody; camelid; anorectic; heavy chain variable domain; VH;  
KW human dietary enzyme inhibitor; medicament; human pancreatic lipase; HPL;  
KW food; human gastric lipase; HGL; cosmetic control; body weight.  
XX  
OS Lama sp.  
XX  
FH Key Location/Qualifiers  
FT Region /label= CDR1  
FT /note= "Complementarity determining region 1"  
FT /label= CDR2  
FT /note= "Complementarity determining region 2"  
FT /label= CDR3  
FT /note= "Complementarity determining region 3"  
XX  
XX EP1134231-A1.  
XX  
XX 19-SEP-2001.  
XX  
XX 20-FEB-2001; 2001EP-00200703.  
XX  
XX 14-MAR-2000; 2000EP-00200930.  
XX  
XX (UNIL ) UNILEVER NV.  
XX (UNIL ) UNILEVER PLC.  
XX

PI Bezemer S, Van De Burg M, De Haard JWM, Tareilus E;  
XX WPI; 2001-572718/65.  
XX  
XX New antibody or its fragments for inhibiting human dietary enzymes,  
PT useful for cosmetic control of body weight of human beings, comprises  
PT heavy chain variable domain derived from immunoglobulin naturally devoid  
XX of light chains.  
XX  
PS Example 2; Page 9; 37pp; English.  
XX  
XX The patent discloses antibodies or their fragments comprising a heavy  
CC chain variable domain (VH) derived from an immunoglobulin naturally  
CC devoid of light chains specific for inhibiting human dietary enzymes. The  
CC antibodies of the invention are useful for the preparation of medicaments  
CC or food for inhibiting the activity of one or more human dietary enzymes  
CC especially human pancreatic lipase (HPL) or human gastric lipase (HGL)  
CC which are useful for the cosmetic control of body weight of human beings.  
CC The present peptide sequence is HPL inhibiting VH fragment, HPL #13 from  
CC llama (camelid) species  
XX  
SQ Sequence 129 AA;  
Query Match 77.3%; Score 532; DB 4; Length 129;  
Best Local Similarity 80.9%; Pred. No. 2.4e-37;  
Matches 106; Conservative 6; Mismatches 17; Indels 2; Gaps 1;  
QY 1 QVQLQESGGGLVQAGSLRLSCAASGSISSINVMGFQAPGKRELVASITSGSTNYA 60  
DB 1 QVQLQESGGGLVQAGSLRLSCAASGSISSINVMGFQAPGKRELVASITSGSTNYA 60  
QY 61 DSLKGRFTISRDNKAAVYIQMNNLKPEDTAVYYCNAHITPAGSSNYYGYWGHGRTKTV 120  
DB 61 DSLKGRFTISRDNKAAVYIQMNNLKPEDTAVYYCNAHITPAGSSNYYGYWGHGRTKTV 118  
QY 121 SSEPKTPKPOP 131  
DB 119 SSEPKTPKPOP 129  
RESULT 4  
AAE10557  
ID AAE10557 standard; peptide; 130 AA.  
XX  
XX AAE10557;  
XX  
DT 10-DEC-2001 (first entry)  
XX  
DE HPL inhibiting VH fragment, HPL #19 from llama species.  
XX  
XX llama antibody; camelid; anorectic; heavy chain variable domain; VH;  
KW human dietary enzyme inhibitor; medicament; human pancreatic lipase; HPL;  
XX food; human gastric lipase; HGL; cosmetic control; body weight.  
XX  
OS Lama sp.  
XX  
FH Key Location/Qualifiers  
FT 31..35  
FT /label=CDR1  
FT /note="Complementarity determining region 1"  
FT 50..64  
FT /label=CDR2  
FT /note="Complementarity determining region 2"  
FT 98..110  
FT /label=CDR3  
FT /note="Complementarity determining region 3"  
XX  
XX EPI134231-A1.  
XX  
XX 19-SEP-2001.  
XX  
XX 20-FEB-2001; 2001EP-00200703.  
XX

PR 14-MAR-2000; 2000EP-00200930.  
XX  
XX (UNIT ) UNILEVER NV.  
FA (UNIT ) UNILEVER PLC.  
PA  
PI Bezemer S, Van De Burg M, De Haard JWM, Tareilus E;  
XX WPI; 2001-572718/65.  
XX  
XX New antibody or its fragments for inhibiting human dietary enzymes,  
PT useful for cosmetic control of body weight of human beings, comprises  
PT heavy chain variable domain derived from immunoglobulin naturally devoid  
XX of light chains.  
XX  
PS Example 2; Page 10; 37pp; English.  
XX  
XX The patent discloses antibodies or their fragments comprising a heavy  
CC chain variable domain (VH) derived from an immunoglobulin naturally  
CC devoid of light chains specific for inhibiting human dietary enzymes. The  
CC antibodies of the invention are useful for the preparation of medicaments  
CC or food for inhibiting the activity of one or more human dietary enzymes  
CC especially human pancreatic lipase (HPL) or human gastric lipase (HGL)  
CC which are useful for the cosmetic control of body weight of human beings.  
CC The present peptide sequence is HPL inhibiting VH fragment, HPL #19 from  
CC llama (camelid) species  
XX  
SQ Sequence 130 AA;  
Query Match 76.1%; Score 523.5; DB 4; Length 130;  
Best Local Similarity 76.3%; Pred. No. 1.3e-36;  
Matches 100; Conservative 10; Mismatches 20; Indels 1; Gaps 1;  
QY 1 QVQLQESGGGLVQAGSLRLSCAASGSISSINVMGFQAPGKRELVASITSGSTNYA 60  
DB 1 QVQLQESGGGLVQAGSLRLSCAASGSISSINVMGFQAPGKRELVASITSGSTNYA 60  
QY 61 DSLKGRFTISRDNKAAVYIQMNNLKPEDTAVYYCNAHITPAGSSNYYGYWGHGRTKTV 120  
DB 61 DSLKGRFTISRDNKAAVYIQMNNLKPEDTAVYYCNAHITPAGSSNYYGYWGHGRTKTV 119  
QY 121 SSEPKTPKPOP 131  
DB 120 SSEPKTPKPOP 130  
RESULT 5  
AAE10554  
ID AAE10554 standard; peptide; 130 AA.  
XX  
XX AAE10554;  
XX  
DT 10-DEC-2001 (first entry)  
XX  
DE HPL inhibiting VH fragment, HPL #14 from llama species.  
XX  
XX llama antibody; camelid; anorectic; heavy chain variable domain; VH;  
KW human dietary enzyme inhibitor; medicament; human pancreatic lipase; HPL;  
XX food; human gastric lipase; HGL; cosmetic control; body weight.  
XX  
OS Lama sp.  
XX  
FH Key Location/Qualifiers  
FT 31..35  
FT /label=CDR1  
FT /note="Complementarity determining region 1"  
FT 50..64  
FT /label=CDR2  
FT /note="Complementarity determining region 2"  
FT 98..110  
FT /label=CDR3  
FT /note="Complementarity determining region 3"  
XX  
XX EPI134231-A1.  
XX

```

XX 19-SEP-2001.
PD 20-FEB-2001; 2001EP-00200703.
XX
XX 14-MAR-2000; 2000EP-00200930.
PR
XX (UNIL ) UNILEVER NV.
XX (UNIL ) UNILEVER PLC.
XX
PI Bezemer S, Van De Burg M, De Haard JWM, Tareilus E;
XX
XX WPI; 2001-572718/65.
XX
XX New antibody or its fragments for inhibiting human dietary enzymes,
XX useful for cosmetic control of body weight of human beings, comprises
XX heavy chain variable domain derived from immunoglobulin naturally devoid
XX of light chains.
XX
XX Example 2; Page 9; 37pp; English.
XX
XX The patent discloses antibodies or their fragments comprising a heavy
XX chain variable domain (VH) derived from an immunoglobulin naturally
XX devoid of light chains specific for inhibiting human dietary enzymes. The
XX antibodies of the invention are useful for the preparation of medicaments
XX or food for inhibiting the activity of one or more human dietary enzymes
XX especially human pancreatic lipase (HPL) or human gastric lipase (HGL)
XX which are useful for the cosmetic control of body weight of human beings.
XX The present peptide sequence is HPL inhibiting VH fragment, HPL #14 from
XX llama (camelid) species
XX
SQ Sequence 130 AA;
Query Match 75.7%; Score 520.5; DB 4; Length 130;
Best Local Similarity 76.3%; Pred. No. 2.3e-36;
Matches 100; Conservative 12; Mismatches 18; Indels 1; Gaps 1;

QY 1 QVQLQESGGGLVQAGGSLRLSCAASGSISSINVMGFRAAPGKORELIVASITSGGSTNYA 60
DB 1 QVQLQESGGGLVQAGGSLRLSCAASGSISSINVMGFRAAPGKORELIVASITSGGSTNYA 60
QY 61 DSLKGRFTISRDNKAVLYQMNNLKPEDTAVVYCNAAHTTPAGSSNYVYGVGHGTRVTV 120
DB 61 DSLKGRFTISRDNKAVLYQMNNLKPEDTAVVYCNAAHTTPAGSSNYVYGVGHGTRVTV 119
QY 121 SSSPKTPKPOP 131
DB 120 SSSPKTPKPOP 130

RESULT 6
AAE10558
ID AAE10558 standard; peptide; 129 AA.
XX
XX AAE10558;
XX
XX 10-DEC-2001 (first entry)
XX
XX HPL inhibiting VH fragment, HPL #22 from llama species.
XX
XX llama antibody; camelid; anorectic; heavy chain variable domain; VH;
XX human dietary enzyme inhibitor; medicament; human pancreatic lipase; HPL;
XX food; human gastric lipase; HGL; cosmetic control; body weight.
XX
XX Lama sp.
XX
XX Key Location/Qualifiers
XX Region 31..35
XX /label= CDR1
XX /note= "Complementarity determining region 1"
XX 50..64
XX /label= CDR2
XX /note= "Complementarity determining region 2"

```

```

FT Region 98..109
FT /label= CDR3
FT /note= "Complementarity determining region 3"
XX
XX EPI134231-A1.
XX
XX 19-SEP-2001.
XX
XX 20-FEB-2001; 2001EP-00200703.
XX
XX 14-MAR-2000; 2000EP-00200930.
XX
XX (UNIL ) UNILEVER NV.
XX (UNIL ) UNILEVER PLC.
XX
XX Bezemer S, Van De Burg M, De Haard JWM, Tareilus E;
XX
XX WPI; 2001-572718/65.
XX
XX New antibody or its fragments for inhibiting human dietary enzymes,
XX useful for cosmetic control of body weight of human beings, comprises
XX heavy chain variable domain derived from immunoglobulin naturally devoid
XX of light chains.
XX
XX Example 2; Page 10; 37pp; English.
XX
XX The patent discloses antibodies or their fragments comprising a heavy
XX chain variable domain (VH) derived from an immunoglobulin naturally
XX devoid of light chains specific for inhibiting human dietary enzymes. The
XX antibodies of the invention are useful for the preparation of medicaments
XX or food for inhibiting the activity of one or more human dietary enzymes
XX especially human pancreatic lipase (HPL) or human gastric lipase (HGL)
XX which are useful for the cosmetic control of body weight of human beings.
XX The present peptide sequence is HPL inhibiting VH fragment, HPL #22 from
XX llama (camelid) species
XX
SQ Sequence 129 AA;
Query Match 73.8%; Score 508; DB 4; Length 129;
Best Local Similarity 77.1%; Pred. No. 2.6e-35;
Matches 101; Conservative 11; Mismatches 17; Indels 2; Gaps 1;

QY 1 QVQLQESGGGLVQAGGSLRLSCAASGSISSINVMGFRAAPGKORELIVASITSGGSTNYA 60
DB 1 QVQLQESGGGLVQAGGSLRLSCAASGSISSINVMGFRAAPGKORELIVASITSGGSTNYA 60
QY 61 DSLKGRFTISRDNKAVLYQMNNLKPEDTAVVYCNAAHTTPAGSSNYVYGVGHGTRVTV 120
DB 61 DSLKGRFTISRDNKAVLYQMNNLKPEDTAVVYCNAAHTTPAGSSNYVYGVGHGTRVTV 118
QY 121 SSSPKTPKPOP 131
DB 119 SSSPKTPKPOP 129

RESULT 7
AAE10559
ID AAE10559 standard; peptide; 130 AA.
XX
XX AAE10559;
XX
XX 10-DEC-2001 (first entry)
XX
XX HPL inhibiting VH fragment, HPL #30 from llama species.
XX
XX llama antibody; camelid; anorectic; heavy chain variable domain; VH;
XX human dietary enzyme inhibitor; medicament; human pancreatic lipase; HPL;
XX food; human gastric lipase; HGL; cosmetic control; body weight.
XX
XX Lama sp.
XX
XX Key Location/Qualifiers
XX Region 31..35

```

FT	/label= CDR1
FT	/note= "Complementarity determining region 1"
FT	49..64
Region	/label= CDR2
FT	/note= "Complementarity determining region 2"
FT	98..110
Region	/label= CDR3
FT	/note= "Complementarity determining region 3"
XX	
PN	EP1134231-A1.
XX	
PD	19-SEP-2001.
XX	
PF	20-FEB-2001; 2001EP-00200703.
PR	14-MAR-2000; 2000EP-00200930.
PA	(UNITL ) UNILEVER NV.
PA	(UNIL ) UNILEVER PLC.
PI	Bezemer S, Van De Burg M, De Haard JWM, Tareilus E;
PS	WPI, 2001-572718/55.
PT	New antibody or its fragments for inhibiting human dietary enzymes,
PT	useful for cosmetic control of body weight of human beings, comprises
PT	heavy chain variable domain derived from immunoglobulin naturally devoid
XX	of light chains.
XX	
BS	Example 2; Page 10; 37pp; English.
CC	The patent discloses antibodies or their fragments comprising a heavy
CC	chain variable domain (VH) derived from an immunoglobulin naturally
CC	devoid of light chains specific for inhibiting human dietary enzymes. The
CC	antibodies of the invention are useful for the preparation of medicaments
CC	or food for inhibiting the activity of one or more human dietary enzymes
CC	specially human pancreatic lipase (HPL) or human gastric lipase (HGL)
CC	which are useful for the cosmetic control of body weight of human beings.
CC	The present peptide sequence is HPL inhibiting VH fragment, HPL #30 from
XX	llama (camelid) species
SQ	Sequence 130 AA:
	Query Match 73.5%; Score 505.5; DB 4; Length 130;
	Best Local Similarity 74.8%; Pred.No.4,2e-35;
	Matches 98; Conservative 11; Mismatches 21; Indels 1; Gaps 1
DY	1 QVQLPDSGGSLVAGGSLRLSCAASGISISINVMGMFROAPRGKOREIVASITSGSTNYA 60
Db	1 QVQLPDSGGSLVAGGSLRLSCAASGISIDITHMGMYRQTGKORDVAVATIGSGSTNYA 60
DY	61 DSLKGRFTIRDNAAKAAYFLQMNNLKPEPTANYVCNAHTTLAGSSNYYGWGHGTNTY 120
Db	61 DSYVGRTTIRDMTLNTVTYLQMWDLPEPTGYWNADVRRPYTSRYL-LTWGGGLTVTV 119
DY	121 SSEPKTPKPOP 131
Db	120 FLEPKTKPKOP 130
RESULT 8	
ABG30620	ABG30620 standard; protein; 131 AA.
AC	ABG30620;
DT	21-OCT-2002 (first entry)
DE	Immunoglobulin G specific heavy chain variable domain antibody #3.
KM	Heavy chain variable domain; antibody; protein array; aging; VHH;
KM	immunoglobulin; sticky phase; antibody; micro-panning; immune library;
KM	proteomic; mouse.

XX	Mus sp.
OS	WO200248193-A2.
PN	20-JUN-2002.
XX	
PD	
PF	03-DEC-2001; 2001WO-EP014471.
XX	
PR	13-DEC-2000; 2000EP-00311142.
XX	
PA	(UNIL ) UNILEVER NV.
FA	(UNIL ) UNILEVER PIC.
XX	(UNIL ) HINDUSTAN LEVER LTD.
P1	De Haard JWW, Hermans P, Landa I, Verrips CT;
DR	WPI; 2002-583487/62.
XX	
PT	Novel protein array useful for detecting the presence of individual
XX	proteins in sample, comprises heavy-chain variable domain antibodies or
PT	antibody fragments obtainable from Camelidae.
XX	
PS	Disclosure; Fig 3; 80pp; English.
CC	The invention describes a protein array (I) comprising a number of heavy-
CC	chain variable domain antibodies or antibody fragments, obtainable from
CC	Camelidae. The method is useful for removing abundant proteins from an
CC	extract or sample which do not provide useful information on the
CC	condition of a cell or tissue in the extract or sample to be
.CC	investigated. (I) is useful for detecting the presence of individual
CC	proteins in a sample, comparing the distribution of proteins in different
CC	cell types, and identification of proteins that may be of importance in
CC	determining the altered properties of cells in disease, aging or other
CC	conditions. Using a heavy-chain variable domain derived from an
CC	immunoglobulin that is naturally devoid of light chains (VHH) in (I)
CC	provides a number of advantages, such as an improvement of
CC	sensitivity/resolution in the order of 10-100 times, and detection of
CC	post-translationally modified proteins. The invention also describes a
CC	method (II) that enables the simultaneous processing of large numbers of
CC	target antigens in a controlled way. The incorporated phage-ELISA
CC	generates on-line information about the success or failure of a certain
CC	panning condition. This feature combined with the microtiter plate format
CC	allows the complete automation of the technology, based on computer-made
CC	decisions on the values of the phage-ELISA for continuation of a limited
CC	number of selections. In (II), due to the fact that many different
CC	conditions can be tested, varying amounts of input-phages can be used
CC	simultaneously in order to decrease the enrichment of sticky phage-
CC	antibodies. Micro-panning is an effective tool for selecting both naive,
CC	synthetic and immune libraries on large numbers of different target
CC	molecules, enabling the generation of large panels of antibodies in short
CC	time frames needed for the generation of arrays (proteincoxs). This
CC	sequence represents an immunoglobulin (IgG) heavy chain variable domain
XX	antibody (VHH)
SQ	Sequence 131 AA;
Query Match	73.4%; Score 505; DB 5; Length 131;
Best Local Similarity	74.8%; Pred. No. 4.7e-35;
Matches	98; Conservative 6; Mismatches 27; Indels 0; Gaps 0
OY	1 QVOLOESGGGLVQAGSGIRLSCAASGSISINWMGFQAPEKQKEIVASITSGGSTNVA 60     :
DB	1 QVOLDSGGGLVQPGSIRLSCAASKSIIFGCGAVGMHRCAPEKQKEIVARITYDGSNTNA 60     :
OY	61 DSLKRFITSRNADNAVYLQGNINKPEDTAYYYCNAHITPAGSSNNYVGWGHGTIKTV 120     :
DB	61 DSVKGRFITSRNADKVLYLQNSLKPEDTYGYVCNAEIVRATIGAFINDLMGGGTIVV 120     :
OY	121 SSEPKTPKPQP 131
DB	121 SSEPKTPKPQP 131

RESULT 9  
AAE10556  
ID AAE10556 standard; peptide; 130 AA.  
XX  
XX AAE10556;  
AC  
XX  
XX 10-DEC-2001 (first entry)  
DT  
XX  
XX HPL inhibiting VHH fragment, HPL #18 from llama species.  
DE  
XX  
XX Llama antibody; camelid; anorectic; heavy chain variable domain; VHH;  
KW human dietary enzyme inhibitor; medicament; human pancreatic lipase; HPL;  
KW food; human gastric lipase; HGL; cosmetic control; body weight.  
XX  
XX Lama sp.  
OS  
XX  
XX Key Location/Qualifiers  
FH 31..35  
FT Region /label= CDR1  
FT /note= "Complementarity determining region 1"  
FT 50..64  
FT Region /label= CDR2  
FT /note= "Complementarity determining region 2"  
FT 98..110  
FT Region /label= CDR3  
FT /note= "Complementarity determining region 3"  
PN  
XX EPI134231-A1.  
XX  
XX 19-SEP-2001.  
PD  
XX 20-FEB-2001; 2001EP-00200703.  
FE  
XX 14-MAR-2000; 2000EP-00200930.  
PR  
XX (UNIL ) UNILEVER NV.  
XX (UNIL ) UNILEVER PLC.  
PA  
XX Bezemer S, Van De Burg M, De Haard JTW, Tareilus E;  
PI WPI; 2001-572718/65.  
PT  
XX New antibody or its fragments for inhibiting human dietary enzymes;  
PT useful for cosmetic control of body weight of human beings, comprises  
PT heavy chain variable domain derived from immunoglobulin naturally devoid  
PT of light chains.  
PS  
XX Example 2; Page 10; 37pp; English.  
XX  
XX The patent discloses antibodies or their fragments comprising a heavy  
CC chain variable domain (VHH) derived from an immunoglobulin naturally  
CC devoid of light chains specific for inhibiting human dietary enzymes. The  
CC antibodies of the invention are useful for the preparation of medicaments  
CC or food for inhibiting the activity of one or more human dietary enzymes  
CC especially human pancreatic lipase (HPL) or human gastric lipase (HGL)  
CC which are useful for the cosmetic control of body weight of human beings.  
CC The present peptide sequence is HPL inhibiting VHH fragment, HPL #18 from  
CC llama (camelid) species  
XX  
XX Sequence 130 AA;  
SQ  
Query Match 73.0%; Score 502.5; DB 4; Length 130;  
Best Local Similarity 76.3%; Pred. No. 7,6e-35;  
Matches 100; Conservative 7; Mismatches 23; Indels 1; Gaps 1;  
QY  
XX 1 QVQLQSGGGGLVQAGGSLRLSCAASGSISSINVMGFRQAPGKQRELVASITSGSTNYA 60  
DB 1 QVQLQDSGGGLVQAGGSLRLSCAASGSISSINVMGFRQAPGKQRELVASITSGSTNYA 60  
QY 61 DSLIKGFTISPDNKAQAVYLOMNNLKPEPTAVYYCNNAHTTPAGSSNYVYGVWGCHGTXTV 120  
pb 61 DPKVGRFTISPDNGKLTIVLOMNSLKPEPTAVYYCNALIRKFTISY-NEIWGGGTQTV 119

QY 121 SSEPTKPKPP 131  
DB 120 SSEPTKPKPP 130  
RESULT 10  
AAE05283  
ID AAE05283 standard; protein; 152 AA.  
XX  
XX AAE05283;  
AC  
XX  
XX 18-SEP-2001 (first entry)  
DT  
XX  
XX Anti-potato SBEII (Clone46) VH region attached with myc and his6 tag.  
DE  
XX  
XX Potato; heavy chain immunoglobulin; pathogen resistance;  
KW metabolism modulator; passive immunisation; heavy chain variable domain;  
KW VH; anti-potato SBEII; starch branching enzyme; SBE A.  
XX  
XX Solanum tuberosum.  
OS Unidentified.  
OS Chimeric.  
OS  
XX EPI118669-A2.  
XX  
XX 25-JUL-2001.  
PD  
XX 08-DEC-2000; 2000EP-00310997.  
PE  
XX 17-DEC-1999; 99EP-00310188.  
FR  
XX (UNIL ) UNILEVER PLC.  
XX (UNIL ) UNILEVER NV.  
PA  
XX Franken LGJ, Van Der Logt CPE, Jobling SA, Teh Y;  
PI WPI; 2001-427157/46.  
XX N-PSDB; AAD10054.  
DR  
XX  
XX Modifying a plant to produce an antibody useful for increasing pathogen  
PT resistance or to modulate metabolism comprises introducing a DNA sequence  
PT encoding a heavy chain immunoglobulin linked to a peptide that targets a  
PT cellular compartment.  
PS  
XX Example 1; Fig 12; 81pp; English.  
XX  
XX The present invention relates to a method for modifying a plant to  
CC produce an antibody or an active fragment or derivative, or a protein  
CC functional equivalent, in a cellular compartment comprises introducing a  
CC DNA sequence encoding a heavy chain immunoglobulin, where the DNA is  
CC linked to promoters and provided with an additional sequence encoding a  
CC peptide capable of targeting heavy chain immunoglobulin to a cellular  
CC compartment. The method is used for producing a heavy chain  
CC immunoglobulin or an active fragment or derivative, or a protein that is  
CC functionally equivalent for increasing the pathogen resistance in a plant  
CC or to modulate metabolism in a plant. Under some circumstances it may be  
CC desirable to retain the antibody product with the plant rather than  
CC extracting and isolating the product. In particular, edible selected  
CC antigens may be used in a method of passively immunising an animal,  
CC preferably human, against the antigen, e.g., pathogenic organisms. The  
CC present sequence is an anti-potato SBEII antibody (denoted Clone46) heavy  
CC chain variable domain (VH) attached to peptide linkers, myc and his6 tag.  
XX The potato SBEII is a starch branching enzyme also designated SBE A  
SQ  
XX Sequence 152 AA;  
SQ  
Query Match 72.9%; Score 501.5; DB 4; Length 152;  
Best Local Similarity 77.1%; Pred. No. 1.1e-34;  
Matches 101; Conservative 7; Mismatches 18; Indels 5; Gaps 1;  
QY  
XX 1 QVQLQSGGGGLVQAGGSLRLSCAASGSISSINVMGFRQAPGKQRELVASITSGSTNYA 60  
pb 1 QVQLQDSGGGLVQAGGSLRLSCAASGSISSINVMGFRQAPGKQRELVASITSGSTNYA 60





CC llama (camelid) species  
 XX Sequence 129 AA;  
 SQ Query March 72.4%; Score 498; DB 4; Length 129;  
 Best Local Similarity 73.5%; Pred. No. 1.8e-34;  
 Matches 97; Conservative 12; Mismatches 19; Indels 4; Gaps 2;

QY 1 QVQLQESGGGIVQAGGSLRLSCAASGSISSINVMGFRQAPGKRELVASITSGSTNYA 60  
 Db 1 QVQLQESGGGIVQAGGSLRLSCAASGSISSINVMGFRQAPGKRELVASITSGSTNYA 60  
 QY 61 DSLKGRFTISRDNAKNSLYLQMNNLKPEDTAVYYCAHITPAGSSNRY-VYGVGHTKRT 119  
 Db 61 DSLKGRFTISRDNAKNSLYLQMNNLKPEDTAVYYCKGR--GSLTQYSEHDWGGSTQVT 117  
 QY 120 VSSEPKTPKPOP 131  
 Db 118 VSSEPKTPKPOP 129

RESULT 13  
 AAE05288 standard; protein; 149 AA.  
 ID AAE05288;  
 AC AAE05288;  
 XX 18-SEP-2001 (first entry)  
 DT Anti-potato SBEII (Clone68) VH region attached with myc and his6 tag.  
 DE Potato; heavy chain immunoglobulin; pathogen resistance;  
 KM metabolism modulator; passive immunisation; heavy chain variable domain;  
 KW VH; anti-potato SBEII; starch branching enzyme; SBE A.  
 OS Solanum tuberosum.  
 OS Unidentified.  
 OS Chimeric.  
 XX EP118669-A2.  
 XX 25-JUL-2001.  
 PD 08-DEC-2000; 2000EP-00310997.  
 PF 17-DEC-1999; 99EP-00310188.  
 PR (UNITL) UNILEVER PLC.  
 PA (UNITL) UNILEVER NV.  
 XX Frenken LGJ, Van Der Logt CPE, Jobling SA, Teh Y;  
 DR WPI; 2001-427157/46.  
 DR N-PSDB; AAD10059.  
 XX PT Modifying a plant to produce an antibody useful for increasing pathogen  
 PT resistance or to modulate metabolism comprises introducing a DNA sequence  
 PT encoding a heavy chain immunoglobulin linked to a peptide that targets a  
 PT cellular compartment.  
 XX Example 13; Fig 28; 81pp; English.

CC The present invention relates to a method for modifying a plant to  
 CC produce an antibody or an active fragment or derivative, or a protein  
 CC functional equivalent, in a cellular compartment comprising introducing a  
 CC DNA sequence encoding a heavy chain immunoglobulin, where the DNA is  
 CC linked to promoters and provided with an additional sequence encoding a  
 CC peptide capable of targeting heavy chain immunoglobulin to a cellular  
 CC compartment. The method is used for producing a heavy chain  
 CC immunoglobulin or an active fragment or derivative, or a protein that is  
 CC functionally equivalent for increasing the pathogen resistance in a plant  
 CC or to modulate metabolism in a plant. Under some circumstances it may be  
 CC desirable to retain the antibody product with the plant rather than

CC extracting and isolating the product. In particular, edible selected  
 CC antigens may be used in a method of passively immunising an animal,  
 CC preferably human, against the antigen, e.g., pathogenic organisms. The  
 CC present sequence is an anti-potato SBEII antibody (denoted Clone68) heavy  
 CC chain variable domain (VH) attached to peptide linkers, myc and his6 tag.  
 CC The potato SBEII is a starch branching enzyme also designated SBE A  
 XX SQ Sequence 149 AA;  
 Query March 72.2%; Score 497; DB 4; Length 149;  
 Best Local Similarity 76.3%; Pred. No. 2.6e-34;  
 Matches 100; Conservative 9; Mismatches 12; Indels 10; Gaps 2;

QY 1 QVQLQESGGGIVQAGGSLRLSCAASGSISSINVMGFRQAPGKRELVASITSGSTNYA 60  
 Db 3 QVQLQESGGGIVQAGGSLRLSCAASGSISSINVMGFRQAPGKRELVASITSGSTNYA 62  
 QY 61 DSLKGRFTISRDNAKNSLYLQMNNLKPEDTAVYYCAHITPAGSSNRY-VYGVGHTKRT 120  
 Db 63 DSLKGRFTISRDNAKNSLYLQMNNLKPEDTAVYYCT---AGGS---YWGQITQVT 112  
 QY 121 SSEPKTPKPOP 131  
 Db 113 SSEPKTPKPOP 123

RESULT 14  
 AAE10563 standard; peptide; 124 AA.  
 ID AAE10563;  
 AC AAE10563;  
 XX 10-DEC-2001 (first entry)  
 DT HGL inhibiting VH fragment, HGL #9 from llama species.  
 DE HGL inhibiting VH fragment, HGL #9 from llama species.  
 XX llama antibody; camelid; anorectic; heavy chain variable domain; VH;  
 KW human dietary enzyme inhibitor; medicament; human pancreatic lipase; HPL;  
 KW food; human gastric lipase; HGL; cosmetic control; body weight.  
 OS Lama sp.  
 OS Location/Qualifiers  
 FH Key 31..35  
 FT Region /label=CDR1  
 FT /note="Complementarity determining region 1"  
 FT Region 50..64  
 FT /label=CDR2  
 FT /note="Complementarity determining region 2"  
 FT Region 98..104  
 FT /label=CDR3  
 FT /note="Complementarity determining region 3"  
 XX EP1134231-A1.  
 XX 19-SEP-2001.  
 XX 20-FEB-2001; 2001EP-00200703.  
 XX 14-MAR-2000; 2000EP-00200930.  
 XX (UNITL) UNILEVER NV.  
 XX (UNITL) UNILEVER PLC.  
 XX Bezemer S, Van De Burg M, De Haard JW, Tareilus E;  
 DR WPI; 2001-572718/65.  
 XX PT New antibody or its fragments for inhibiting human dietary enzymes,  
 PT useful for cosmetic control of body weight of human beings, comprises  
 PT heavy chain variable domain derived from immunoglobulin naturally devoid  
 PT of light chains.  
 XX

PS Example 4; Page 13; 37pp; English.

XX The patent discloses antibodies or their fragments comprising a heavy

CC chain variable domain (VH) derived from an immunoglobulin naturally

CC devoid of light chains specific for inhibiting human dietary enzymes. The

CC antibodies of the invention are useful for the preparation of medicaments

CC or food for inhibiting the activity of one or more human dietary enzymes

CC especially human pancreatic lipase (HPL) or human gastric lipase (HGL)

CC which are useful for the cosmetic control of body weight of human beings.

CC The present peptide sequence is HGL inhibiting VH fragment, HGL #9 from

CC llama (camelid) species

XX

SQ Sequence 124 AA;

Query Match 72.2%; Score 496.5; DB 4; Length 124;

Best Local Similarity 74.0%; Pred. No. 2.3e-34;

Matches 97; Conservative 10; Mismatches 17; Indels 7; Gaps 2;

QY 1 QVQLQESGGGVQAGSRLSCAASGSISSINVMGFQAPGKQRELVAISITSGSTNYA 60

DB 1 QVQLQESGGGVQAGSRLSCAASGSISSINVMGFQAPGKQRELVAISITSGSTNYA 60

QY 61 DSLKGRFTISRDNKNAKAVYLQMNILKPEDTAVYVCNAHITPAGSSNVYVYGWGHGKTKYV 120

DB 61 DSVKGRFTISRDNKNTWYLQMNLSLKPEDTGYVC---AGTGAAGH---YWGQGTQVTV 113

QY 121 SSEPKTKPKPOP 131

DB 114 SSEPKTKPKPOP 124

RESULT 15

AAE10564

ID AAE10564 standard, peptide; 129 AA.

XX

AAE10564;

XX

DT 10-DEC-2001 (first entry)

XX

DE HGL inhibiting VH fragment, HGL #10 from llama species.

XX

KW llama antibody; camelid; anorectic; heavy chain variable domain; VH;

KW human dietary enzyme inhibitor; medicament; human pancreatic lipase; HPL;

KW food; human gastric lipase; HGL; cosmetic control; body weight.

XX

OS Lama sp.

XX

Key Location/Qualifiers

FT 31..35

FT Region

FT /label= CDR1

FT /note= "Complementarity determining region 1"

FT 50..64

FT /label= CDR2

FT /note= "Complementarity determining region 2"

FT 95..109

FT /label= CDR3

FT /note= "Complementarity determining region 3"

XX

PN EP1134231-A1.

XX

PD 19-SEP-2001.

XX

PF 20-FEB-2001; 2001EP-00200703.

XX

PR 14-MAR-2000; 2000EP-00200930.

XX

PA (UNITL ) UNILEVER NV.

PA (UNITL ) UNILEVER PLC.

PI Bezemer S, Van De Burg W, De Haard JWM, Tareilus E;

XX WPI; 2001-572718/65.

XX

PT New antibody or its fragments for inhibiting human dietary enzymes,

PT useful for cosmetic control of body weight of human beings, comprises

PT heavy chain variable domain derived from immunoglobulin naturally devoid

PT of light chains.

XX

PS Example 4; Page 13; 37pp; English.

XX

CC The patent discloses antibodies or their fragments comprising a heavy

CC chain variable domain (VH) derived from an immunoglobulin naturally

CC devoid of light chains specific for inhibiting human dietary enzymes. The

CC antibodies of the invention are useful for the preparation of medicaments

CC or food for inhibiting the activity of one or more human dietary enzymes

CC especially human pancreatic lipase (HPL) or human gastric lipase (HGL)

CC which are useful for the cosmetic control of body weight of human beings.

CC The present peptide sequence is HGL inhibiting VH fragment, HGL #10 from

CC llama (camelid) species

XX

SQ Sequence 129 AA;

Query Match 71.8%; Score 494; DB 4; Length 129;

Best Local Similarity 73.5%; Pred. No. 3.9e-34;

Matches 97; Conservative 11; Mismatches 20; Indels 4; Gaps 2;

QY 1 QVQLQESGGGVQAGSRLSCAASGSISSINVMGFQAPGKQRELVAISITSGSTNYA 60

DB 1 QVQLQESGGGVQAGSRLSCAASGSISSINVMGFQAPGKQRELVAISITSGSTNYA 60

QY 61 DSLKGRFTISRDNKNAKAVYLQMNILKPEDTAVYVCNAHITPAGSSNVYVYGWGHGKTKYV 119

DB 61 NSVKGRTISRDNKNTWYLQMNLSLKPEDTAVYVCGR---GGLTGYSEHDVWGQGTQVTV 117

QY 120 VSSEPKTKPKPOP 131

DB 118 VSSEPKTKPKPOP 129

Search completed: October 5, 2004, 08:00:11

Job time : 99.6651 secs

This Page Blank (uspio)

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd

OM protein - protein search, using sw model

Run on: October 5, 2004, 07:36:16 ; Search time 21.0648 Seconds  
(without alignments)  
593.639 Million cell updates/sec

Title: US-09-805-290A-23  
 Perfect score: 685  
 Sequence: 1 QVQLQDSGGGLVQAGSLRL...WGCGTQVTVSSEPKTPKQP 130

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

```
Post-processing: Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries
```

```
Database : PIR_78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

Result	Query	Match	DB	ID	Description
No.	Score	Length			
1	381.5	55.7	120	2	1g heavy chain V H
2	371	54.2	119	2	1g heavy chain - I
3	371	54.2	121	2	1g heavy chain - I
4	371	54.2	123	2	1g heavy chain - I
5	369	53.9	119	2	1g heavy chain - I
6	368	53.7	160	2	1g heavy chain pre
7	362	52.8	117	2	1g heavy chain - I
8	362	52.8	119	2	1g heavy chain V I
9	360.5	52.6	114	2	1g heavy chain V I
10	359	52.4	140	2	1g heavy chain V I
11	357	52.1	119	2	1g heavy chain V I
12	357	52.1	125	2	1g heavy chain V I
13	357	52.1	140	2	1g heavy chain V I
14	356	52.0	138	2	1g heavy chain - I
15	355.5	51.9	114	2	1g heavy chain - I
16	355	51.8	135	2	1g heavy chain V I
17	353.5	51.6	114	2	1g heavy chain V I
18	353.5	51.6	140	2	1g heavy chain pre
19	353	51.5	121	2	1g heavy chain V I
20	352	51.4	119	2	1g heavy chain V I
21	352	51.4	121	2	1g heavy chain V I
22	351.5	51.3	136	2	1g heavy chain V I
23	351.5	51.3	147	2	1g heavy chain V I
24	351	51.2	120	2	1g heavy chain V I
25	351	51.2	120	2	1g heavy chain V I
26	351	51.2	134	2	1g heavy chain V I
27	351	51.2	139	2	1g heavy chain V I
28	350.5	51.2	120	2	1g heavy chain V I
29	350.5	51.2	124	2	1g heavy chain V I

30	350.5	51.2	128	2	S25786	Ig heavy chain V
31	350	51.1	127	2	S27486	Ig heavy chain V
32	350	51.1	123	2	A36706	Ig heavy chain V
33	350	51.1	123	2	S26794	Ig heavy chain V
34	349.5	51.0	143	2	S23624	Ig heavy chain V
35	349	50.9	121	2	S19666	Ig heavy chain V
36	349	50.9	121	2	S31106	Ig heavy chain V
37	349	50.9	132	2	S31603	Ig heavy chain V
38	348.5	50.9	118	2	S31105	Ig heavy chain (su
39	348.5	50.9	122	2	S31117	Ig heavy chain - h
40	348	50.8	139	2	I37781	Ig variable region
41	347.5	50.7	114	2	S46392	Ig heavy chain V
42	347.5	50.7	116	2	S31110	Ig heavy chain - h
43	347.5	50.7	133	2	S31610	Ig heavy chain - h
44	347.5	50.7	141	2	S31669	Ig heavy chain V
45	347	50.7	121	2	S25798	Ig heavy chain V

## ALIGNMENTS

## RESULT 1

Ig heavy chain V region (anti-Sm, VH3/Dxg4/CH4b) - human  
C:Species: Homo sapiens (man)  
C:Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 23-Jul-1999  
C:Accession: S48798  
R:Mamoudi, M.; Edwards, J.; Cairns, E.; Bell, D.  
submitted to the EMBL Data Library, October 1994  
A:Description: Molecular characterization of natural human anti-Sm autoantibodies  
A:Reference number: S48797  
A:Accession: S48798  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-120 <MMH>  
A:Cross-references: EMBL:Z46382; NID:G562324; PID:CAA6521.1; PID:G1340167  
C:Superfamily: immunoglobulin V region; immunoglobulin homology  
C:Keywords: heterotetramer; immunoglobulin  
F:15-98/Domain: immunoglobulin homology <IMM>

Query Match	55.7%	Score 381.5	DB 2	Length 120
Best Local Similarity	67.2%	Pred. No. 4.6e-26		
Matches 82	Conservative 9	Mismatches 28	Indels 3	Gaps 2

[illegible]

## RESULT 2

I9 heavy chain human  
 C|Species: Homo sapiens (man)  
 C|Date: 02-Dec-1993 #sequence\_revision 26-May-1995 #ext\_change 17-Mar-1999  
 C|Accession: S31107  
 R|Raphoport, F.M., Timmers, E., Kenter, M.T.H., van Tol, M.J.D., Vossen, J.M., Schuurman  
 E|ur, J. Immunol. 22, 247-251, 1992  
 A|Title: Restricted utilization of germ-line V(H)3 genes and short diverse third compleme  
 A|Reference number: S31104; NID:92111633; PMID:1730252  
 A|Accession: S31107  
 A|Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A|Molecule type: mRNA  
 A|Residues: 1-119 <RA>  
 A|Cross-references: EMBL:X62955  
 A|Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1991

C:Superfamily: immunoglobulin V region; immunoglobulin homology  
C:Keywords: heterotetramer; immunoglobulin  
F:15-98/Domain: immunoglobulin homology <IMM>

Query Match 54.2%; Score 371; DB 2; Length 119;

Best Local Similarity 65.6%; Pred. No. 3.7e-25;  
Matches 80; Conservative 10; Mismatches 28; Indels 4; Gaps 2;

1 QVQLQDSGGGLVQAGGSLRLSCAASGTTIDITMAWHROAPGKERELVASATES-GSPNY 59  
1 EVQLVESGGGLVQPGGSLRLSCAASGFTFSVAMSWVRQAPGKLEWVSATISGSGSTYY 60

60 ADPVKGRFTISRDNKLTLYLQNNSLKPEDTAVYYCNALIRKFTSEVNEVWGCGTQVTV 119

61 ADSVKGRFTISRDNKNTLYLQNNSLRAEDTAVYYCAAKPAHAGSPY-DYWGQGLTVTV 117

QY 120 SS 121  
DB 118 SS 119

## RESULT 3

155673  
Ig heavy chain - human (fragment)

C:Species: Homo sapiens (man)  
C>Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 21-Jan-2000

C:Accession: I55673

R:Knights, G.B.; Agnello, V.; Bonagura, V.; Barnes, J.L.; Panka, D.J.; Zhang, Q.X.  
J. Exp. Med. 178, 1903-1911, 1993

A>Title: Human rheumatoid factor cross-idiotypes. IV. Studies on WA Xid-positive IGM with  
tinct from the 17.109 and G6 Xids.

A:Reference number: I55673; MUID:94065558; PMID:8245772

A:Accession: I55673

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-121 <RES>

A:Cross-references: GB:M87268; NID:g186197; PIDN:AAC37536.1; PID:g186198

C:Genetics:  
A:Gene: GDB:IGHM

A:Cross-references: GDB:120086; OMIM:147020

A:Map position: 14q32.33-14q32.33

C:Superfamily: immunoglobulin V region; immunoglobulin homology  
F:15-98/Domain: immunoglobulin homology <IMM>

Query Match 54.2%; Score 371; DB 2; Length 121;  
Best Local Similarity 65.6%; Pred. No. 3.8e-25;

Matches 80; Conservative 10; Mismatches 30; Indels 2; Gaps 2;

1 QVQLQDSGGGLVQAGGSLRLSCAASGTTIDITMAWHROAPGKERELVASATES-GSPNY 59  
1 EVQLVESGGGLVQPGGSLRLSCAASGFTFSVAMSWVRQAPGKLEWVSATISGSGSTYY 60

60 ADPVKGRFTISRDNKLTLYLQNNSLKPEDTAVYYCNALIRKFTSEVNEVWGCGTQVTV 119

61 ADSVKGRFTISRDNKNTLYLQNNSLRAEDTAVYYCAAPPAHAGSPY-DYWGQGLTVTV 119

QY 120 SS 121  
DB 120 SS 121

RESULT 4  
S3114  
Ig heavy chain - human

C:Species: Homo sapiens (man)  
C>Date: 02-Dec-1993 #sequence\_revision 26-May-1995 #text\_change 17-Mar-1999

C:Accession: S3114  
R:Radphoret, F.M.; Timmers, E.; Kenter, M.J.H.; van Tol, M.J.D.; Vossen, J.M.; Schuurman  
Eur. J. Immunol. 22, 247-251, 1992

A>Title: Restricted utilization of germ-line V(H)3 genes and short diverse third comple  
A:Reference number: S31104; MUID:92111633; PMID:1730252

A:Accession: S31114  
A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: mRNA  
A:Residues: 1-123 <RAA>

A:Cross-references: EMBL:X62963

A>Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1991

C:Superfamily: immunoglobulin V region; immunoglobulin homology  
C:Keywords: heterotetramer; immunoglobulin  
F:15-98/Domain: immunoglobulin homology <IMM>

Query Match 54.2%; Score 371; DB 2; Length 123;  
Best Local Similarity 65.0%; Pred. No. 3.8e-25;

Matches 80; Conservative 11; Mismatches 30; Indels 2; Gaps 2;

1 QVQLQDSGGGLVQAGGSLRLSCAASGTTIDITMAWHROAPGKERELVASATES-GSPNY 59  
1 EVQLVESGGGLVQPGGSLRLSCAASGFTFSVAMSWVRQAPGKLEWVSATISGSGSTYY 60

60 ADPVKGRFTISRDNKLTLYLQNNSLKPEDTAVYYCNALIRKFTSEVNEVWGCGTQVTV 118

61 ADSVKGRFTISRDNKNTLYLQNNSLRAEDTAVYYCAKSLYRFLFEMFLDYWGQGLTVTV 120

QY 119 VSS 121  
DB 121 VSS 123

## RESULT 5

S31108  
Ig heavy chain - human

C:Species: Homo sapiens (man)  
C>Date: 02-Dec-1993 #sequence\_revision 26-May-1995 #text\_change 17-Mar-1999

C:Accession: S31108

R:Radphoret, F.M.; Timmers, E.; Kenter, M.J.H.; van Tol, M.J.D.; Vossen, J.M.; Schuurman,  
Eur. J. Immunol. 22, 247-251, 1992

A>Title: Restricted utilization of germ-line V(H)3 genes and short diverse third comple  
A:Reference number: S31104; MUID:92111633; PMID:1730252

A:Accession: S31108

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: mRNA

A:Residues: 1-119 <RAA>

A:Cross-references: EMBL:X62956

A>Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1991

C:Superfamily: immunoglobulin V region; immunoglobulin homology  
C:Keywords: heterotetramer; immunoglobulin  
F:15-98/Domain: immunoglobulin homology <IMM>

Query Match 53.9%; Score 369; DB 2; Length 119;  
Best Local Similarity 65.6%; Pred. No. 5.5e-25;

Matches 80; Conservative 11; Mismatches 27; Indels 4; Gaps 3;

1 QVQLQDSGGGLVQAGGSLRLSCAASGTTIDITMAWHROAPGKERELVASATES-GSPNY 59  
1 EVQLVESGGGLVQPGGSLRLSCAASGFTFSVAMSWVRQAPGKLEWVSATISGSGSTYY 60

60 ADPVKGRFTISRDNKLTLYLQNNSLKPEDTAVYYCNALIRKFTSEVNEVWGCGTQVTV 119

61 ADSVKGRFTISRDNKNTLYLQNNSLRAEDTAVYYCAK--DRPLGTFL-DYWGQGLTVTV 117

QY 120 SS 121  
DB 118 SS 119

RESULT 6  
S05271  
Ig heavy chain precursor - human (fragment)

C:Species: Homo sapiens (man)  
C>Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 16-Aug-1996

C:Accession: S05271; S04602  
R:Kishimoto, T.  
submitted to the EMBL Data Library, March 1989

A:Reference number: S05270  
A:Accession: S05271  
A:Molecule type: mRNA

A:Residues: 1-160 <K151>  
A:Cross-references: EMBL:X14584  
R:Kishimoto, T.; Okajima, H.; Okumoto, T.; Taniguchi, M.  
Nucleic Acids Res. 17, 4385, 1989  
A>Title: Nucleotide sequences of the cDNAs encoding the V-regions of H- and L-chains of  
A:Reference number: S04601; MUID:89296497; PMID:2500644  
A:Accession: S04602  
A:Molecule type: mRNA  
A:Residues: 1-144 <K152>  
A:Cross-references: EMBL:X14584  
C:Superfamily: immunoglobulin V region; immunoglobulin homology  
C:Keywords: heterodimer; immunoglobulin  
F:1-19/Domain: signal sequence status predicted <Sig>  
F:20-160/Product: Ig heavy chain (fragment) #status predicted <Mat>  
F:34-117/Domain: immunoglobulin homology <Im>

Query Match	53.7%;	Score 368;	DB 2;	Length 160;
Best Local Similarity	64.0%;	Pred. No. 9.2e-25;		
Matches	80;	Conservative 10;	Mismatches 31;	Indels 4;
				Gaps 2;

Qy	1	OVLCPDGGGGLVAGGSLRISCAASGTIDITYAMHMQAPKXKEELVAASITEE--GSPNY	59
Db	20	EVLITBEGGGLVPGGSRLRISCAASGFSTSYAAMSWRQAPCKGLEWVASISGSGSNTY	79
Qy	60	ADPYKGPFTTIRNGKLTVYLQNNLSLKPEDTAIVYYCNALIRXKF <sup>a</sup> SEY---NEWVGQTQ	116
Db	80	ADSVKGRFTTIRSDNSKRTLYLQNNLSLRADTAVYYCAARVARGVISTYYYEMDMWGQTT	139
Qy	117	VTVSS	121
Db	140	VTVSS	144

```

RESULT 7
S31109
Ig heavy chain - human
C:Species: Homo sapiens (man)
C:Date: 02-Dec-1993 #sequence_revision 26-May-1995 #text_change 17-Mar-1999
C:Accession: S31109
C:Raphoport, F.M.; Timmers, E.; Kentler, M.J.H.; van Toi, M.J.D.; Voessen, J.M.; Schuurman
Eur. J. Immunol. 22, 247-251, 1992
A:Title: Restricted utilization of germ-line V(H)3 genes and short diverse third comple
A:Reference number: S31104; MUID:92111633; PMID:11730252
A:Accession: S31109
A:Status: Preliminary; nucleic acid sequence not shown; translation not shown
A:Stetus: Preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: mRNA
A:Residues: 1-117 <RAA>
A:Cross-references: EMBL:X62960
A>Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1991
C:Superfamily: immunoglobulin V region; immunoglobulin homology
C:Keywords: heterotetramer; immunoglobulin
C:15-100/Domain: immunoglobulin homology <IMM>

```

Query Match 52.8%; Score 362; DB 2; Length 117;  
Best Local Similarity 62.9%; Pred. No. 2-2e-24;  
Matches 78; Conservative 11; Mismatches 25; Indels 10; Gaps 2

```

QY      |OVLQDSDGGGLVQAGSLRFLSCASGFTIGIYMAHMRQAPKREELVA---SATESSP 57
        |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      |EVLQVBSGGGLVKGGSRLRSCASGFTSNMAMSVVRQAPKRGLEWGRISKIDGTT 60
        |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY      |58 NYADPVKGRFTISRDNCKLTVYIQMNSLRPEADPAVYVYCAALIRKFTSENYMGCGTQV 11
        |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      |61 DYADPVKGRFTISRDKSKNTLLYIQMNSLKIEDIAYVYCA-----TYFDYMGCGTLV 11
        |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY      |118 TVSS 121
        |:::|:::|
Db      |114 TVSS 117

```

RESULT 8  
D36005  
Ig heavy chain V region (M43) - human

```

C|Species: Homo sapiens (man)
C|Date: 21-Dec-1990 #sequence_revision 21-Dec-1990 #ext_change 16-Dec-1998
C|Accession: D36005
R|Schroeder Jr., H.W.; Wang, J.Y.
Proc. Natl. Acad. Sci. U.S.A. 87, 6146-6150, 1990
A|Title: Preferential utilization of conserved immunoglobulin heavy chain variable gene
A|Reference number: A36005; MUID:90349571; PMID:2117273
A|Accession: D36005
A|Status: preliminary
A|Molecule type: mRNA
A|Residues: 1-119 <SCCH>
A|Cross-references: GB:M34024
C|Genetics:
A|Gene: GDB:IGH@; IGHDIY1
A|Cross-references: GDB:118731; OMIM:146910
A|Map position: 14q32.33-14q32.33
C|Superfamily: immunoglobulin V region; immunoglobulin homology
C|Keywords: heterotrimer; immunoglobulin
E|15-98/Domain: immunoglobulin homology <IMM>

```

Query March	52.8%;	Score 362;	DB 2;	Length 119;
Best Local Similarity	63.1%;	Pred. No. 2.2e-24;		
Matches 77;	Conservative 13;	Mismatches 28;	Indels 4;	Gaps 2

[illegible]

RESULT 9  
S46391  
IG heavy chain V region - human  
C.Species: Homo sapiens (man)  
C.Date: 27-Jan-1995 #sequence\_revision 27-Jan-1995 #next\_change 20-Jun-2000  
C.Accession: S46391  
R.Figini, M.; Marks, J.D.; Winter, G.; Griffiths, A.D.  
J. Mol. Biol. 239, 68-78, 1994  
A.Title: In vitro assembly of repertoires of antibody chains on the surface of phage by  
A.Reference number: S46390; MUID:94254092; PMID:8196048  
A.Accession: S46391  
A.Status: preliminary  
A.Molecule type: DNA  
A.Residues: 1-114 <FIG>  
A.Cross-references: EXBL:Z31687; NID:G509784; PIDN:CA83492.1; PID:G1335144  
C.Superfamily: immunoglobulin V region; immunoglobulin homology  
C.Keywords: heterotetramer; immunoglobulin  
F15-98/Domain: immunoglobulin homology <Imm>

Query Match	52.6%;	Score 360.5;	DB 2;	Length 114;
Best Local Similarity	63.1%;	Pred. No. 2.8e-24;		
Matches	77;	Conservative	8;	Mismatches 28;
				Indels 9;
				Gaps 2;

QY 1 QVQLDPSGSGGLVWAGSSTLSLSCAASGTIGDIITAMNHQAQKEELIASATIEGSP-Y 5  
:::  
Db 1 QVNLRGGGSLVPGGSILRLSCASGFPTSSVAHWVRQAQGKLEWAVALVIWDGSKYY 60

QY 60 ADPKSGFTSRNGKLTVYLQMNSLKPEDTAVYYCNALIRKFTSEVENYAGQQTIV 119  
:::  
Db 61 ADSYKGRFTSRNKSNTLYLQMSLRADDTAVYYC-----ARDGDITWGGLTIV 112

QY 120 SS 121  
:::  
Db 113 SS 114

## RESULT 10

S31588

Ig heavy chain V region - human (fragment)

C/Species: Homo sapiens (man)

C/Date: 22-Nov-1993 #sequence\_revision 10-Nov-1995 #text\_change 23-Jul-1999

C/Accession: S31588

R/Cuisinier, A.M.; Gautier, L.; Boubli, L.; Fougereau, M.; Tonnelle, C.

submitted to the EMBL Data Library, June 1992

A/Description: Mechanisms that generate human immunoglobulin diversity operate from the

A/Reference number: S31585

A/Accession: S31588

A/Status: preliminary

A/Molecule type: mRNA

A/Residues: 1-140 &lt;CUI&gt;

A/Cross-references: EMBL:Z14200; NID:G30957; PIDN:CAA7869.1; PID:G30958

C/Superfamily: immunoglobulin V region; immunoglobulin homology

C/Keywords: heterotetramer; immunoglobulin

F/34-117/Domain: immunoglobulin homology &lt;IMM&gt;

Query Match 52.1%; Score 357; DB 2; Length 140;

Best Local Similarity 61.2%; Pred. No. 4.8e-24;

Matches 79; Conservative 11; Mismatches 23; Indels 16; Gaps 3;

QY 1 QVQLDSSGGGLVQAGSLRLSCAASGTTIDITMAHWRQAPGKERELVASATES-GSPNY 59

DB 20 EVQLVESGGGLVPGGSLRLSCAASGFTFSYAMSVWRQAPGKLEWVSAISGSGSTYY 79

QY 60 ADPVKGRFTISRDNKLTLYLQNSLKPEDTAVYYCNALIRKFTSEYNEWVG 112

DB 80 ADSVKGKFTISRDNKLTLYLQNSLRADTAVYYC-AKDRNDSSGYSHYFDWGQGT 131

QY 113 QGTQTVSS 121

DB 132 QGTQTVSS 140

## RESULT 11

C36005

Ig heavy chain V region (30p1) - human

C/Species: Homo sapiens (man)

C/Date: 21-Dec-1990 #sequence\_revision 21-Dec-1990 #text\_change 16-Aug-1996

C/Accession: C36005

R/Schroeder Jr., H.W.; Wang, J.Y.

Proc. Natl. Acad. Sci. U.S.A. 87, 6146-6150, 1990

A/Title: Preferential utilization of conserved immunoglobulin heavy chain variable gene

A/Reference number: A36005; MID:90349571; PMID:2117273

A/Accession: C36005

A/Status: preliminary

A/Molecule type: mRNA

A/Residues: 1-119 &lt;SCH&gt;

A/Cross-references: GB:M18513

C/Superfamily: immunoglobulin V region; immunoglobulin homology

C/Keywords: heterotetramer; immunoglobulin

F/15-98/Domain: immunoglobulin homology &lt;IMM&gt;

Query Match 52.1%; Score 357; DB 2; Length 119;

Best Local Similarity 63.9%; Pred. No. 6e-24;

Matches 78; Conservative 10; Mismatches 30; Indels 4; Gaps 2;

QY 1 QVQLDSSGGGLVQAGSLRLSCAASGTTIDITMAHWRQAPGKERELVASATES-GSPNY 59

DB 1 EVQLVESGGGLVPGGSLRLSCAASGFTFSYAMSVWRQAPGKLEWVSAISGSGSTYY 60

QY 60 ADPVKGRFTISRDNKLTLYLQNSLKPEDTAVYYCNALIRKFTSEYNEWVGQGTQTV 119

DB 61 ADSVKGKFTISRDNKLTLYLQNSLRADTAVYYC-AKDAWGSGFDYWGQGTQTV 117

QY 120 SS 121

DB 118 SS 119

## RESULT 12

S30531

Ig heavy chain V region - human

C/Species: Homo sapiens (man)

C/Date: 06-Jan-1995 #sequence\_revision 06-Jan-1995 #text\_change 16-Aug-1996

C/Accession: S30531

R/Mariette, X.

submitted to the EMBL Data Library, October 1992

A/Reference number: S30520

A/Accession: S30531

A/Status: preliminary

A/Molecule type: mRNA

A/Residues: 1-125 &lt;MAR&gt;

A/Cross-references: EMBL:Z18317

C/Superfamily: immunoglobulin V region; immunoglobulin homology

C/Keywords: heterotetramer; immunoglobulin

F/15-98/Domain: immunoglobulin homology &lt;IMM&gt;

Query Match 52.1%; Score 357; DB 2; Length 125;

Best Local Similarity 62.7%; Pred. No. 6.3e-24;

Matches 79; Conservative 11; Mismatches 30; Indels 6; Gaps 3;

QY 1 QVQLDSSGGGLVQAGSLRLSCAASGTTIDITMAHWRQAPGKERELVASATESGSP-NY 59

DB 1 EVQLVESGGGLVPGGSLRLSCAASGFTFSYAMSVWRQAPGKLEWVSAISGSGSTYY 60

QY 60 ADPVKGRFTISRDNKLTLYLQNSLKPEDTAVYYCNALIRKFTSEYNEWVGQGT 115

DB 61 ADSVKGKFTISRDNKLTLYLQNSLRADTAVYYC-ASRNVDSSGYSHYFDWGQGT 119

QY 116 QTVSS 121

DB 120 LTVSS 125

## RESULT 13

S31686

Ig heavy chain V region - human (fragment)

C/Species: Homo sapiens (man)

C/Date: 22-Nov-1993 #sequence\_revision 10-Nov-1995 #text\_change 23-Jul-1999

C/Accession: S31686

R/Cuisinier, A.M.; Gautier, L.; Boubli, L.; Fougereau, M.; Tonnelle, C.

submitted to the EMBL Data Library, June 1992

A/Description: Mechanisms that generate human immunoglobulin diversity operate from the

A/Reference number: S31685

A/Accession: S31686

A/Status: preliminary

A/Molecule type: mRNA

A/Residues: 1-140 &lt;CUI&gt;

A/Cross-references: EMBL:Z14205; NID:G30969; PIDN:CAA78574.1; PID:G30970

C/Superfamily: immunoglobulin V region; immunoglobulin homology

C/Keywords: heterotetramer; immunoglobulin

F/34-117/Domain: immunoglobulin homology &lt;IMM&gt;

Query Match 52.1%; Score 357; DB 2; Length 140;

Best Local Similarity 63.7%; Pred. No. 7.1e-24;

Matches 79; Conservative 11; Mismatches 28; Indels 6; Gaps 3;

QY 1 QVQLDSSGGGLVQAGSLRLSCAASGTTIDITMAHWRQAPGKERELVASATES-GSPNY 59

DB 20 EVQLVESGGGLVPGGSLRLSCAASGFTFSYAMSVWRQAPGKLEWVSAISGSGSTYY 79

QY 60 ADPVKGRFTISRDNKLTLYLQNSLKPEDTAVYYCNALIRKFTSEYNEWVGQGTQTV 117

DB 80 ADSVKGKFTISRDNKLTLYLQNSLRADTAVYYC-AKCPFAGSGSFDYWGQGTQTV 136

QY 118 TVSS 121

DB 137 TVSS 140

## RESULT 14

S31666

Ig heavy chain V region - human (fragment)





This Page Blank (uspto)

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2004, 07:29:35 ; Search time 11.6358 Seconds  
(Without alignments)  
581.749 Million cell updates/sec

Title: US-09-805-290A-23

Perfect score: 685  
Sequence: 1 QVQLDSGGGLVQAGGSURL.....WGQGTQVTVSSSEPKPKPPQ 130

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	342.5	50.0	126	1	HV3K_HUMAN
2	336	49.1	119	1	HV3T_HUMAN
3	329	48.0	115	1	HV3D_HUMAN
4	327.5	47.8	116	1	HV3T_HUMAN
5	325.5	47.5	122	1	HV3G_HUMAN
6	324.5	47.4	114	1	HV01_CANFA
7	324	47.3	115	1	HV3F_HUMAN
8	318.5	46.5	122	1	HV3A_HUMAN
9	317.5	46.4	119	1	HV40_MOUSE
10	317	46.3	121	1	HV3J_HUMAN
11	316.5	46.2	117	1	HV3C_HUMAN
12	313.5	45.8	120	1	HV3U_HUMAN
13	313	45.7	119	1	HV3L_HUMAN
14	313	45.7	120	1	HV3E_HUMAN
15	312.5	45.6	114	1	HV3B_HUMAN
16	312.5	45.6	136	1	HV16_MOUSE
17	310	45.3	115	1	HV33_MOUSE
18	309	45.1	116	1	HV05_CARAU
19	308.5	45.0	119	1	HV37_MOUSE
20	306	44.7	115	1	HV32_MOUSE
21	305	44.5	113	1	HV30_MOUSE
22	304	44.4	117	1	HV02_CANFA
23	302.5	44.2	122	1	HV20_MOUSE
24	301.5	44.0	117	1	HV30_HUMAN
25	301.5	44.0	122	1	HV3H_HUMAN
26	301	43.9	113	1	HV27_MOUSE
27	300.5	43.9	119	1	HV3P_HUMAN
28	300.5	43.9	119	1	HV3P_HUMAN
29	299	43.6	118	1	HV39_MOUSE
30	298.5	43.6	119	1	HV38_MOUSE
31	298.5	43.6	119	1	HV3X_HUMAN
32	298.5	43.6	122	1	HV21_MOUSE
33	297	43.4	113	1	HV31_MOUSE

34	295.5	43.1	111	1	HV35_MOUSE
35	295	43.1	113	1	HV28_MOUSE
36	295	43.1	123	1	HV18_MOUSE
37	295	43.1	123	1	HV22_MOUSE
38	295	43.1	123	1	HV23_MOUSE
39	295	43.1	142	1	HV01_RAT
40	294	42.9	123	1	HV19_MOUSE
41	292.5	42.7	117	1	HV53_MOUSE
42	292	42.6	97	1	HV56_MOUSE
43	292	42.6	123	1	HV24_MOUSE
44	291.5	42.6	117	1	HV54_MOUSE
45	291	42.5	113	1	HV29_MOUSE

## ALIGNMENTS

RESULT 1  
HV3K\_HUMAN STANDARD; PRT; 126 AA.  
ID HV3K\_HUMAN  
AC P01772;  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Ig heavy chain V-II region KOL.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE AND DISULFIDE BONDS.  
RX MEDLINE=83289131; Pubmed=684994;  
RA Schmidt W.B., Jung H.-D., Palm W., Hilschmann N.;  
RT "Three-dimensional structure determination of antibodies. Primary structure of crystallized monoclonal immunoglobulin IgG1 KOL, I",  
RL Hoppe-Seyler's Z. Physiol. Chem. 364:713-747(1983).  
RN [2]  
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).  
RX MEDLINE=61072295; Pubmed=7441755;  
RA Marguier M., Deisenhofer J., Huber R., Palm W.;  
RT "Crystallographic refinement and atomic models of the intact immunoglobulin molecule KOL and its antigen-binding fragment at 3.0 A and 1.0-A resolution.",  
RL J. Mol. Biol. 141:369-391(1980).  
CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.  
DR PIR; A02055; GIHUKL.  
DR PDB; 2FB4; 12-JUL-89.  
DR GO; GO:0005576; C:extracellular; NAS.  
DR GO; GO:0003823; F:antigen binding; NAS.  
DR GO; GO:0006955; P:immune response; NAS.  
DR InterPro; IPR007110; IG-like.  
DR Pfam; PFO0047; IG\_1.  
DR SMART; SMO0406; IGV\_1.  
DR PROSITE; PS00835; IG\_LIKE; 1.  
KM Immunoglobulin V region; 3D-structure; Pyrrolidone carboxylic acid.  
FT MOA\_RES 1 112  
FT DOA\_RES 1 112  
FT DISULFID 22 96  
FT DISULFID 105 110  
FT STRAND 3 7  
FT STRAND 11 12  
FT TURN 14 15  
FT STRAND 18 15  
FT STRAND 18 25  
FT HELIX 29 31  
FT STRAND 34 39  
FT TURN 41 42  
FT STRAND 45 51  
FT TURN 53 54  
FT STRAND 58 60  
FT HELIX 62 64  
FT STRAND 65 65

FT TURN 66 67  
 FT STRAND 68 73  
 FT TURN 74 77  
 FT STRAND 76 83  
 FT HELIX 88 90  
 FT STRAND 92 99  
 FT TURN 106 106  
 FT STRAND 107 108  
 FT STRAND 109 109  
 FT STRAND 113 116  
 FT STRAND 120 124  
 FT NON\_TER 126 126  
 SQ SEQUENCE 126 AA, 13718 MW, E4D71B52B16F8776 CRC64,

Query Match 50.0%; Score 342.5; DB 1; Length 126;  
 Best Local Similarity 59.5%; Pred. No. 6.5e-26;  
 Matches 75; Conservative 13; Mismatches 33; Indels 5; Gaps 2;

QY 1 QVQLQDSGGGLVQAGSIRLSCAASGTIGDITMAMHROAPGKERELVASATSSG-SPY 59  
 DB 1 QVQLVSGGGVQPGSLRLSCSSGPFSSYAMVWRQAPGKLEWVAITWDDSDQHY 60

QY 60 ADPVKGRFTISRDNKLTLYLQNNSLKPEDTAVYYCNALIRKFTSEYN---EYWGQGT 115  
 DB 61 ADPVKGRFTISRDNKLTLYLQNNSLKPEDTAVYYCNALIRKFTSEYN---EYWGQGT 120

QY 116 QVTWSS 121  
 DB 121 PVTWSS 126

## RESULT 2

ID HV3D HUMAN STANDARD; PRT; 119 AA.  
 AC P01770;

DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Ig heavy chain V-II region NIE.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominda; Homo.  
 OC NCBI\_TaxId=9606;

RN [1]  
 RP SEQUENCE.  
 RX MEDLINE=77070269; PubMed=826475;

RA Penstingl H., Hilschmann N.;  
 RT "The rule of antibody structure. The primary structure of a  
 RT monoclonal IgG1 immunoglobulin (myeloma protein Nle). III. The  
 RT chymotryptic peptides of the H-chain, alignment of the tryptic  
 RT peptides and discussion of the complete structure.";  
 RL Hoppe-Seyler's Z. Physiol. Chem. 357:1571-1604 (1976).

RN [2]  
 RP DISULFIDE BOND.  
 RX MEDLINE=77070267; PubMed=1002129;

RA Dreker L., Schwarz J., Reichel W., Hilschmann N.;  
 RT "Rule of antibody structure. The primary structure of a monoclonal  
 RT IgG1 immunoglobulin (myeloma protein Nle). I. Purification and  
 RT characterization of the protein, the L- and H-chains, the  
 RT cyanoogen bromide cleavage products, and the disulfide bridges.";  
 RL Hoppe-Seyler's Z. Physiol. Chem. 357:1515-1540 (1976).

CC -1- MISCELLANEOUS: THIS CHAIN WAS ISOLATED FROM AN IgG1 MYELOMA  
 CC PROTEIN.

CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.  
 DR PIR; A9168; G1HUN1.

DR HSSP; P01772; 2FB4.  
 DR GO; GO:0005576; C:extracellular; NAS.  
 DR GO; GO:0003823; F:antigen binding; NAS.  
 DR GO; GO:0006953; P:immune response; NAS.  
 DR InterPro; IPR007110; IG-like.  
 DR Pfam; PF00047; IG\_V.  
 DR SMART; SM00406; IG\_V; 1.

DR PROSITE; PS50835; IG LIKE; 1.  
 KW Immunoglobulin V region; Pyroglutamate carboxylic acid.  
 FT DOMAIN 1 112  
 FT MOD\_RES 1 1  
 FT DISULFID 22 96  
 FT NON\_TER 119 119  
 SQ SEQUENCE 119 AA, 13242 MW, C96935A655E165B CRC64;

Query Match 49.1%; Score 336; DB 1; Length 119;  
 Best Local Similarity 60.7%; Pred. No. 2.5e-25;  
 Matches 74; Conservative 12; Mismatches 32; Indels 4; Gaps 2;

QY 1 QVQLQDSGGGLVQAGSIRLSCAASGTIGDITMAMHROAPGKERELVASATSSG-SPY 59  
 DB 1 QVQLVSGGGVQPGSLRLSCSSGPFSSYAMVWRQAPGKLEWVAITWDDSDQHY 60

QY 60 ADPVKGRFTISRDNKLTLYLQNNSLKPEDTAVYYCNALIRKFTSEYNWQGTQVTV 119  
 DB 61 ADPVKGRFTISRDNKLTLYLQNNSLKPEDTAVYYCNALIRKFTSEYNWQGTQVTV 117

QY 120 SS 121  
 DB 118 SS 119

## RESULT 3

ID HV3D HUMAN STANDARD; PRT; 115 AA.  
 AC P01765;

DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Ig heavy chain V-II region TIL.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominda; Homo.  
 OC NCBI\_TaxId=9606;

RN [1]  
 RP SEQUENCE.  
 RX MEDLINE=78005528; PubMed=409716;

RA Wang A.-C., Wang T.Y., Fudenberg H.H.;  
 RT "Immunoglobulin structure and genetics. Identity between variable  
 RT regions of a mu and a gamma2 chain.";  
 RL J. Biol. Chem. 252:7192-7199 (1977).

CC -1- MISCELLANEOUS: THE SEQUENCES OF THE V REGIONS OF THE HEAVY CHAINS  
 CC OF IGM AND IGG2 ISOLATED FROM A SINGLE PATIENT WITH BICLONAL  
 CC GAMOPATHY ARE IDENTICAL. THEIR LIGHT CHAINS ARE APPARENTLY ALSO  
 CC IDENTICAL.

CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.  
 DR PIR; A02048; H3HUT1.

DR HSSP; P01772; 2FB4.  
 DR GO; GO:0005576; C:extracellular; NAS.  
 DR GO; GO:0003823; F:antigen binding; NAS.  
 DR GO; GO:0006953; P:immune response; NAS.  
 DR InterPro; IPR007110; IG-like.  
 DR Pfam; PF00047; IG\_V.

DR SMART; SM00406; IG\_V; 1.  
 DR PROSITE; PS50835; IG LIKE; 1.  
 KW Immunoglobulin V region.  
 FT DOMAIN 1 108  
 FT NON\_TER 115 115  
 SQ SEQUENCE 115 AA, 12356 MW, 4DC67D179F62326 CRC64;

Query Match 48.0%; Score 329; DB 1; Length 115;  
 Best Local Similarity 59.0%; Pred. No. 1.1e-24;  
 Matches 72; Conservative 14; Mismatches 28; Indels 8; Gaps 3;

QY 1 QVQLQDSGGGLVQAGSIRLSCAASGTIGDITMAMHROAPGKERELVASATSSG-SPY 59  
 DB 1 QVQLVSGGGVQPGSLRLSCSSGPFSSYAMVWRQAPGKLEWVAITWDDSDQHY 60

QY 60 ADPVKGRFTISRDNKLTLYLQNNSLKPEDTAVYYCNALIRKFTSEYNWQGTQVTV 119  
 DB 61 ADPVKGRFTISRDNKLTLYLQNNSLKPEDTAVYYCNALIRKFTSEYNWQGTQVTV 117

```

Db      61 ABSYKGRFTISRDBSKNT-----MSLRADPNTAVYIC---AAGKYSAYIFBWSGTLVTV 113
QY      120 SS 121
Db      114 SS 115

```

## RESULT 4

```

HV3T_HUMAN          STANDARD;          PRT;          116 AA.
ID   HV3T_HUMAN
AC   P01781
DT   21-JUL-1986 (Rel. 01, Created)
DR   21-JUL-1986 (Rel. 01, Last sequence update)
DE   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Ig heavy chain V-III region GALT.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE.
RX   MEDLINE=75059123; PubMed=4803843;
RA   Watanabe S., Barikol H.U., Horn J., Berttram J., Hilschmann N.;
RT   "The primary structure of a monoclonal IGM-immunoglobulin
RT   (macroglobulin Gal.); II: the amino acid sequence of the H-chain (mu-
RT   type), subgroup H III. Architecture of the complete IGM-molecule."
RL   Hoppe-Seyler's Z. Physiol. Chem. 354:1505-1509(1973).
RN   [2]
RP   REVISION TO 28-33.
RA   Hilschmann N.;
RL   Submitted (JUN-1975) to the PIR data bank.
CC   -1- MISCELLANEOUS: THIS MU CHAIN WAS ISOLATED FROM A WALDENSTROM'S
CC   MACROGLOBULIN.
CC   -1- SIMILARITY: Contains 1 immunoglobulin-like domain.
DR   HSBP, P01772; 2FB4.
DR   GO; GO:0005576; C:extracellular; NAS.
DR   GO; GO:0003823; F:antigen binding; NAS.
DR   GO; GO:0006955; P:immune response; NAS.
DR   InterPro: IPR007110; IG-Like.
DR   InterPro: IPR003596; IG_v.
DR   Pfam: PF00047; Ig_1.
DR   SMART; SM00406; IGV; 1.
DR   PROSITE; PS00835; IG-LIKE; 1.
KW   Immunoglobulin V region.
FT   DOMAIN 1
FT   NON_TER 116
FT   SEQUENCE 116 AA; 12730 MW; 2667CA9AAAAA1282 CRC64;
SQ

```

```

Query Match      47.8%; Score 327.5; DB 1; Length 116;
Best Local Similarity 59.0%; Pred. No. 1.6e-24;
Matches 72; Conservative 13; Mismatches 30; Indels 7; Gaps 3;
QY      1 OVQLDSSGGGLVQAGSLRLSCAASGTDIDYTMAMHQAQPKKEELVASATESGS-PNY 59
DB      1 EVQLVESGGGLVQPGSRRLSCAASGFBPRBBLGWTWQAQPKGLEWYANIKZBSGZBY 60
QY      60 ADPVKGRFTISRDNKLTLYIQLMNSLKPEDTAVYICNALIRKFTSEVNEYWGQGTQVTV 119
DB      61 VDSVKGKRFITSRDNKNSLYIQLMNSLRVEDTALYYC---ARGWGG--GDYWGQGTTLVTV 114
QY      120 SS 121
Db      115 ST 116

```

```

RESULT 5
HV3G_HUMAN          STANDARD;          PRT;          122 AA.
ID   HV3G_HUMAN
AC   P01768;
DT   21-JUL-1986 (Rel. 01, Created)
DR   21-JUL-1986 (Rel. 01, Last sequence update)

```

```

DE   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Ig heavy chain V-III region CAM.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE.
RX   MEDLINE=81013859; PubMed=6774332;
RA   Lehman D.W., Putnam F.W.;
RT   "Amino acid sequence of the variable region of a human mu chain:
RT   location of a possible VH segment."
RL   Proc. Natl. Acad. Sci. U.S.A. 77:3239-3243(1980).
CC   -1- MISCELLANEOUS: THIS MU CHAIN WAS ISOLATED FROM THE PLASMA OF A
CC   PATIENT WITH MACROGLOBULINEMIA.
CC   -1- SIMILARITY: Contains 1 immunoglobulin-like domain.
DR   PIR; A02051; M3HUM.
DR   HSBP, P01772; 2FB4.
DR   GO; GO:0005576; C:extracellular; NAS.
DR   GO; GO:0003823; F:antigen binding; NAS.
DR   GO; GO:0006955; P:immune response; NAS.
DR   InterPro: IPR007110; IG-Like.
DR   InterPro: IPR003596; IG_v.
DR   Pfam; PF00047; Ig_1.
DR   SMART; SM00406; IGV; 1.
DR   PROSITE; PS00835; IG-LIKE; 1.
KW   Immunoglobulin V region; Pyroglutone carboxylic acid.
FT   DOMAIN 1
FT   MOD_RES 112
FT   NON_TER 1
FT   SEQUENCE 122 AA; 13668 MW; A42D0F17D252F1C2 CRC64;
SQ

```

```

Query Match      47.5%; Score 325.5; DB 1; Length 122;
Best Local Similarity 58.1%; Pred. No. 2.6e-24;
Matches 72; Conservative 11; Mismatches 36; Indels 5; Gaps 3;
QY      1 OVQLDSSGGGLVQAGSLRLSCAASGTDIDYTMAMHQAQPKKEELVASATESGS-PNY 59
DB      1 QVELVESGGGVVPGGSLRLSCAASGFTPSNVMHWVROPGGGLEWAVISYGBBKTY 60
QY      60 ADPVKGRFTISRDNKLTLYIQLMNSLKPEDTAVYICNALIRKFTSEYN--EYWGQGTQV 117
DB      61 ABSYKGRFTISRDBSKNTLYIQLMNSLRADPNTAVYIC--DRLYGBYAFNFWGQGTLY 118
QY      118 TVSS 121
Db      119 TVSS 122

```

## RESULT 6

```

HV01_CANFA          STANDARD;          PRT;          114 AA.
ID   HV01_CANFA
AC   P01784;
DT   21-JUL-1986 (Rel. 01, Created)
DR   21-JUL-1986 (Rel. 01, Last sequence update)
DE   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Ig heavy chain V region GOM.
OS   Canis familiaris (Dog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX   NCBI_TaxID=9615;
RN   [1]
RP   SEQUENCE.
RX   MEDLINE=77242268; PubMed=407924;
RA   Wasserman R.L., Capra J.D.;
RT   "Primary structure of the variable regions of two canine
RT   immunoglobulin heavy chains."
RL   Biochemistry 16:3160-3168(1977).
CC   -1- MISCELLANEOUS: THIS CHAIN WAS ISOLATED FROM A MYELOMA PROTEIN.
CC   -1- SIMILARITY: Contains 1 immunoglobulin-like domain.
DR   PIR; A02067; AYDGM.
DR   HSBP, P01772; 2FB4.
DR   InterPro: IPR007110; IG-Like.

```

RESULT 9
HV40_MOUSE
ID HV40_MOUSE STANDARD; PRT; 119 AA.
AC P01810;
DT 21-JUL-1986 (Rel. 01; Created)
DT 21-JUL-1986 (Rel. 01; Last sequence update)
DT 28-FEB-2003 (Rel. 01; Last annotation update)
DE heavy chain V region J539.
MS Mus musculus (Mouse).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi

CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OK NCBI\_TaxId=10090;  
 RN [1]  
 RP PRELIMINARY SEQUENCE.  
 RX MEDLINE=79223895; PubMed=11245;  
 RA Rao D.N., Rudikoff S., Krutzsch H., Potter M.;  
 RT "Structural evidence for independent joining region gene in  
 immunoglobulin heavy chains from anti-galactan myeloma proteins and  
 its potential role in generating diversity in  
 complementarity-determining regions.";  
 RT Proc. Natl. Acad. Sci. U.S.A. 76:2890-2894(1979).  
 RN [2]  
 RP X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS).  
 RX MEDLINE=88217852; PubMed=3449853;  
 RA Sun S.W., Bhat T.N., Navia M.A., Cohen G.H., Rao D.N., Rudikoff S.,  
 RA Davies D.R.;  
 RT "The galactan-binding immunoglobulin Fab U539: an X-ray diffraction  
 study at 2.6-A resolution.";  
 RT Proteins 1:74-80(1986).  
 CC -I MISCELLANEOUS: THIS CHAIN WAS ISOLATED FROM A MYELOMA PROTEIN THAT  
 BINDS GALACTAN.  
 CC PIR: A02080; AVMSJ5.  
 DR PDB: 2FBJ; 15-OCT-90.  
 DR InterPro: IPR007110; IG\_1-like.  
 DR InterPro: IPR003596; IG\_V.  
 DR Pfam: PF00047; IG\_1.  
 DR SMART: SM00406; IGv\_1.  
 DR PROSITE: PS50835; IG\_LIKE; 1.  
 KW Immunoglobulin V region; 3D-structure.  
 FT NON\_TER 119  
 FT STRAND 3  
 FT STRAND 7  
 FT STRAND 10  
 FT TURN 12  
 FT STRAND 14  
 FT STRAND 15  
 FT STRAND 16  
 FT STRAND 25  
 FT HELIX 29  
 FT STRAND 31  
 FT STRAND 34  
 FT TURN 39  
 FT STRAND 41  
 FT STRAND 42  
 FT STRAND 45  
 FT TURN 51  
 FT STRAND 53  
 FT TURN 54  
 FT STRAND 58  
 FT TURN 60  
 FT STRAND 62  
 FT TURN 67  
 FT STRAND 72  
 FT STRAND 78  
 FT STRAND 83  
 FT HELIX 88  
 FT STRAND 90  
 FT STRAND 92  
 FT TURN 100  
 FT TURN 101  
 FT STRAND 103  
 FT STRAND 104  
 FT STRAND 108  
 FT STRAND 112  
 FT STRAND 116  
 SQ SEQUENCE 119 AA; 13240 MW; 577B4F1DB675C1F1 CRC64;

Query Match 46.4%; Score 317.5; DB 1; Length 119;  
 Best Local Similarity 54.5%; Pred. No. 1.5e-23;  
 Matches 67; Conservative 17; Mismatches 34; Indels 5; Gaps 2;

QY 1 QVQLDSDGGGVQAGGSLRLSCAASGTTGDIYTMANHRQAPGKERLVAATESGSPN-Y-59  
 Db 1 EVKLLESGGGLVPGGSLRLSCAASGTPSPSKYMKSWVQAPGKGLIEWLGEIHPDSGTNY-60  
 QY 60 ADPVKGRFTISRDNGKLTIVYIQMNSLKPEDTAVYVCNALLIRKFTSEYNEYWGQGTQVTV-119  
 Db 61 TPLSLDKFTISRDNKNSIYLQMSKVRSHDTALYYC-----ARLHYGYNAVWGQGTIVTV-116  
 QY 120 SSE 122  
 Db 117 SAE 119  
 RESULT 10  
 ID HV3J\_HUMAN STANDARD; PRT; 121 AA.  
 AC P01771;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Ig heavy chain V-II region H1L.  
 OS Homo sapiens (human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 OK NCBI\_TaxId=9606;  
 RN [1]  
 RP SEQUENCE.  
 RX MEDLINE=79124695; PubMed=420800;  
 RA Chiu Y.-X.H., Lopes de Castro J.A., Poljak R.J.;  
 RT "Amino acid sequence of the VH region of human myeloma  
 RT cryoimmunoglobulin IgS H1.";  
 RT Biochemistry 18:553-560(1979).  
 CC -I MISCELLANEOUS: THIS CHAIN WAS ISOLATED FROM AN IGG1 MYELOMA  
 CC PROTEIN.  
 CC -I SIMILARITY: Contains 1 immunoglobulin-like domain.  
 CC PIR: A02054; GHUHL.  
 DR HSSP: P01772; 2FBJ.  
 DR GO: GO:0005576; Extracellular; NAS.  
 DR GO: GO:0003823; Antigen binding; NAS.  
 DR GO: GO:0005955; P:immune response; NAS.  
 DR InterPro: IPR007110; IG\_1-like.  
 DR InterPro: IPR003596; IG\_V.  
 DR SMART: PF00047; IG\_1.  
 DR SMART: SM00406; IGv\_1.  
 DR PROSITE: PS50835; IG\_LIKE; 1.  
 KW Immunoglobulin V region; Pyrolydine carboxylic acid.  
 FT DOMAIN 1 112  
 FT MOD\_RES 1 1  
 FT NON\_TER 121 121  
 FT STRAND 121  
 FT STRAND 121  
 SQ SEQUENCE 121 AA; 13566 MW; 480FC53610E5DAB CRC64;

Query Match 46.3%; Score 317; DB 1; Length 121;  
 Best Local Similarity 57.1%; Pred. No. 1.7e-23;  
 Matches 72; Conservative 10; Mismatches 34; Indels 10; Gaps 3;

QY 1 QVQLDSDGGGVQAGGSLRLSCAASGTTGDIYTMANHRQAPGKERLVAATESGSPN-Y-59  
 Db 1 EVKLLESGGGLVPGGSLRLSCAASGTPSPSKYMKSWVQAPGKGLIEWLGEIHPDSGTNY-60  
 QY 60 ADPVKGRFTISRDNGKLTIVYIQMNSLKPEDTAVYVCNALLIRKFTSEYNEYWGQGTQVTV-115  
 Db 61 GDSVKGKFTISRDNKNSIYLQMSKVRSHDTALYYC-----DIMGQV-115  
 QY 116 QVTSS 121  
 Db 116 LVTSS 121

RESULT 11  
 ID HV3C\_HUMAN STANDARD; PRT; 117 AA.  
 AC P01764;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Ig heavy chain V-II region VH26 precursor.  
 OS Homo sapiens (human).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 OK NCBI\_TaxId=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=8101090; PubMed=6450418;  
 RA Mathysens G., Rabbits T.H.;  
 RT "Structure and multiplicity of genes for the human immunoglobulin  
 RT heavy chain variable region.";  
 RT Proc. Natl. Acad. Sci. U.S.A. 77:6561-6565(1980).  
 CC -I SIMILARITY: Contains 1 immunoglobulin-like domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: J00236; AA53516.1; -  
 DR EMBL: M35415; AA58735.1; -  
 DR PIR: A02047; H3H026.  
 DR PDB: 1HOU; 23-DEC-99.  
 DR Genew; HGNC:5545; IGHV@.  
 DR GO; GO:0005576; C:extracellular; NAS.  
 DR GO; GO:0003823; P:antigen binding; NAS.  
 DR GO; GO:0006955; P:immune response; NAS.  
 DR InterPro; IPR007110; IG-like.  
 DR InterPro; IPR003596; IG\_V.  
 DR SMART; SMART0406; IGV; 1.  
 DR SMART; SMART0407; IGV; 1.  
 DR PROSITE; PSS0835; IG\_LIKE; 1.  
 KM Immunoglobulin V region; Signal, 3D-structure.  
 FT STGNL 1 19  
 FT CHAIN 20 117 IG HEAVY CHAIN V-III REGION VH26.  
 FT DOMAIN 20 >117 IG-LIKE.  
 FT NON\_TER 117 117  
 SQ SEQUENCE 117 AA; 12582 MW; E826733F1A3C80F1 CRC64;

Query Match 46.2%; Score 316.5; DB 1; Length 117;  
 Best Local Similarity 68.8%; Pred. No. 18e-23;  
 Matches 66; Conservative 8; Mismatches 21; Indels 1; Gaps 1;

QY 1 QVQLDQSGGGLVQAGGSLRLSCAASGTTIDITMAMHROAPGKERELVASATES-GSPNY 59  
 DB 20 EVQLTSSGGGLVQPGGSLRLSCAASGFTPSVAMSVRQAPFGKLEWVASLISGSGSTYY 79  
 QY 60 ADPVKGRFTISRPNKLTIVYLCQNSLKPEDTAVYVC 95  
 DB 80 GDSVKGRTISRPNKLTIVYLCQNSLKPEDTAVYVC 115

## RESULT 12

HV3U\_HUMAN STANDARD; PRT; 120 AA.  
 AC P01782;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Ig heavy chain V-III region DOB.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OC NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE.  
 RX MEDLINE=80020921; PubMed=114209;  
 RA Steiner L.A., Garcia Pardo A., Margolis M.N.;  
 RT "Amino acid sequence of the heavy-chain variable region of the  
 RT crystallizable human myeloma protein DOB";  
 RL Biochemistry 18:4068-4080(1979).  
 RN [2]  
 RP CRYSTALLIZATION.  
 RX MEDLINE=80020920; PubMed=114208;  
 RA Steiner L.A., Lopes A.D.;  
 RT "The crystallizable human myeloma protein DOB has a hinge-region  
 RT deletion";  
 RL Biochemistry 18:4054-4067(1979).  
 CC -1- MISCELLANEOUS: THIS GAMMA-1 MYELOMA PROTEIN HAS A DELETION IN THE  
 CC HINGE REGION. THERE ARE NO LIGHT-HEAVY OR INTER-HEAVY CHAIN  
 CC DISULFIDE BONDS.  
 CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.  
 DR PIR: A90431; GIHUB.  
 DR HSSP; P01772; 2FB4.  
 DR GO; GO:0005576; C:extracellular; NAS.  
 DR GO; GO:0003823; P:antigen binding; NAS.  
 DR GO; GO:0006955; P:immune response; NAS.

DR InterPro; IPR007110; IG-like.  
 DR InterPro; IPR003596; IG\_V.  
 DR Pfam; PF00047; IGV; 1.  
 DR SMART; SMART0406; IGV; 1.  
 DR PROSITE; PSS0835; IG\_LIKE; 1.  
 KM Immunoglobulin V region.  
 FT DOMAIN 1 112  
 FT CHAIN 120 120 IG-LIKE.  
 FT NON\_TER 120 120  
 SQ SEQUENCE 120 AA; 13440 MW; 880DE307C4B2627 CRC64;

Query Match 45.8%; Score 313.5; DB 1; Length 120;  
 Best Local Similarity 57.7%; Pred. No. 3.6e-23;  
 Matches 71; Conservative 11; Mismatches 36; Indels 5; Gaps 3;

QY 1 QVQLDQSGGGLVQAGGSLRLSCAASGTTIDITMAMHROAPGKERELVASAT-EGSGPNY 59  
 DB 1 EVQLTSSGGGLVQPGGSLRLSCAASGFTPSVAMSVRQAPFGKLEWVASLISGSGSTYY 60  
 QY 60 ADPVKGRFTISRPNKLTIVYLCQNSLKPEDTAVYVC-NALIRKPTSENEYGQSTQVT 118  
 DB 61 ADSVKGRTISRPNKLTIVYLCQNSLKPEDTAVYVC---NKGWTFDSWGQSTQVT 117  
 QY 119 VSS 121  
 DB 118 VSS 120

## RESULT 13

HV3L\_HUMAN STANDARD; PRT; 119 AA.  
 AC P01773;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Ig heavy chain V-III region BUR.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OC NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE (MYELOMA PROTEIN BUR).  
 RX MEDLINE=79151016; PubMed=107164;  
 RA Putnam F.W., Liu Y.-S.V., Low T.L.K.;  
 RT "Primary structure of a human IgM immunoglobulin. IV. Streptococcal  
 RT IgM protease, digestion, Fab and Fc fragments, and the complete  
 RT amino acid sequence of the alpha 1 heavy chain";  
 RL J. Biol. Chem. 254:2865-2874(1979).  
 CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.  
 DR PIR: A02056; A1HUBR.  
 DR HSSP; P01772; 2FB4.  
 DR GO; GO:0005576; C:extracellular; NAS.  
 DR GO; GO:0003823; P:antigen binding; NAS.  
 DR GO; GO:0006955; P:immune response; NAS.  
 DR InterPro; IPR007110; IG-LIKE.  
 DR InterPro; IPR003596; IG\_V.  
 DR Pfam; PF00047; IGV; 1.  
 DR SMART; SMART0406; IGV; 1.  
 DR PROSITE; PSS0835; IG\_LIKE; 1.  
 KM Immunoglobulin V region; Glycoprotein; Pyroliidone carboxylic acid.  
 FT DOMAIN 1 112  
 FT MOD\_RES 1 112 PYROLIIDONE CARBOXYLIC ACID.  
 FT DISULFID 22 96  
 FT CARBOHYD 28 28 N-LINKED (GLCNAC. . .).  
 FT NON\_TER 119 119  
 SQ SEQUENCE 119 AA; 12981 MW; 12A709A75344D024 CRC64;

Query Match 45.7%; Score 313; DB 1; Length 119;  
 Best Local Similarity 57.9%; Pred. No. 4e-23;  
 Matches 70; Conservative 12; Mismatches 35; Indels 4; Gaps 2;

QY 1 QVQLDQSGGGLVQAGGSLRLSCAASGTTIDITMAMHROAPGKERELVASATESGPNY 59  
 DB 1 QVQLVDSGGGLVQAGGSLRLSCAASGFTPSVAMSVRQAPFGKLEWVASLISGSGSTYY 60



QY 60 ADPVKGRFTSRDNGKLTIVLQNNSLKPEDTAVYYCNALIRKFTSEYNEWGGQTQVTV 119  
 DB 61 ADVSRGRFTISRBSKSTLTLYLZMKTURTREDTAVYYCAALIVAGT---RRRWGGTLVTV 117  
 QY 120 S 120  
 DB 118 S 118

## RESULT 14

HV3B\_HUMAN STANDARD; PRT; 120 AA.  
 ID HV3B\_HUMAN STANDARD; PRT; 120 AA.  
 AC P01766;

DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Ig heavy chain V-III region BFO.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_Taxid=9606;  
 RN [1]  
 RP MEDLINE=77117674; PubMed=65324;

RA Capra J.D., Hopper J.E.  
 RT "Comparative studies on monocytic IgM lambda and IgG kappa from an individual patient. III. The complete amino acid sequence of the VH region of the IgM paraprotein."  
 RL Immunohemistry 13:995-999(1976).  
 CC -I- MISCELLANEOUS: THIS CHAIN WAS OBTAINED FROM IGM ISOLATED FROM THE SERUM OF A PATIENT WITH MALIGNANT LYMPHOMA OF THE WALDENSTROM TYPE.

CC -I- SIMILARITY: Contains 1 immunoglobulin-like domain.  
 DR PIR; A02049; M3HUBW.  
 DR HSSP; P01772; 2FB4.

DR GO; GO:0005576; C:extracellular; NAS.  
 DR GO; GO:0003823; F:antigen binding; NAS.  
 DR GO; GO:0006955; P:immune response; NAS.  
 DR InterPro; IPR007110; Ig-like.  
 DR InterPro; IPR003596; Ig\_V.  
 DR Pfam; PF00047; Ig\_1.  
 DR SMART; SM00406; IGV; 1.  
 DR PROSITE; PS00835; IG\_LIKE; 1.  
 KW Immunoglobulin V region.  
 FT DOMAIN 1 111 IG-LIKE.  
 FT NON\_TER 120 120  
 SQ SEQUENCE 120 AA; 13227 MW; D3F0428F7C2B6410 CRC64;

Query Match 45.7%; Score 313; DB 1; Length 120;  
 Best Local Similarity 56.3%; Pred. No. 4e-23;  
 Matches 67; Conservative 11; Mismatches 37; Indels 4; Gaps 1;

QY 1 QVQLDSSGGGLVQAGSLRLSCAASGTTIGDITVMAMHRQAPKEREVLASATSSGSPNYA 60  
 DB 1 EVQLVESGGGLVQPGGSLRLSCAASGFTFSYNNMWVRQVGTGKLEWVSALGTADQYTA 60  
 QY 61 DSVKGRFTISRBSKSTLTLYLQNNSLKPEDTAVYYCN---ALIRKFTSEYNEWGGQT 115  
 DB 61 DSVKGRFTISRBSKSTLTLYLQNNSLKPEDTAVYYCARSPVSLVDGMVLYYIGSVWGGT 119

## RESULT 15

HV3B\_HUMAN STANDARD; PRT; 114 AA.  
 ID HV3B\_HUMAN STANDARD; PRT; 114 AA.  
 AC P01763;

DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Ig heavy chain V-III region WEA.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

OX NCBI\_Taxid=9606;

RN [1]  
 RP MEDLINE=83273707; PubMed=6410338;  
 RA Goni F., Frangione B.;  
 RT "Amino acid sequence of the Fv region of a human monoclonal IgM galactose in Klebsiella polysaccharides K30 and K33."  
 RL Proc. Natl. Acad. Sci. U.S.A. 80:4837-4841(1983).  
 CC -I- MISCELLANEOUS: THIS CHAIN WAS OBTAINED FROM A MONOCLONAL ANTIBODY AGAINST 3,4-PYRUVYLATED GALACTOSE AND ISOLATED FROM A PATIENT WITH WALDENSTROM'S MACROGLOBULINEMIA.  
 CC -I- SIMILARITY: Contains 1 immunoglobulin-like domain.

DR PIR; A02046; M3HUBW.  
 DR HSSP; P01772; 2FB4.  
 DR GO; GO:0005576; C:extracellular; NAS.  
 DR GO; GO:0003823; F:antigen binding; NAS.  
 DR GO; GO:0006955; P:immune response; NAS.  
 DR InterPro; IPR007110; Ig-like.  
 DR InterPro; IPR003596; Ig\_V.  
 DR Pfam; PF00047; Ig\_1.  
 DR SMART; SM00406; IGV; 1.  
 DR PROSITE; PS00835; IG\_LIKE; 1.  
 KW Immunoglobulin V region; Pyroglutamate carboxylic acid.  
 FT DOMAIN 1 112  
 FT MOD\_RES 1 1 PYROGLUTAMATE CARBOXYLIC ACID.  
 FT NON\_TER 114 114  
 SQ SEQUENCE 114 AA; 12256 MW; D88294FB418A07B7 CRC64;

Query Match 45.6%; Score 312.5; DB 1; Length 114;  
 Best Local Similarity 58.2%; Pred. No. 4.2e-23;  
 Matches 71; Conservative 11; Mismatches 31; Indels 9; Gaps 2;

QY 1 QVQLDSSGGGLVQAGSLRLSCAASGTTIGDITVMAMHRQAPKEREVLASATSSGSP-NY 59  
 DB 1 QVQLVESGGGLVQPGGSLRLSCAASGFTFSANDMNMVRQVGTGKLEWVSALGTADQYTA 60  
 QY 60 ADPVKGRFTSRDNGKLTIVLQNNSLKPEDTAVYYCNALIRKFTSEYNEWGGQTQVTV 119  
 DB 61 ADVSRGRFTISRBSKSTLTLYLQNNSLKPEDTAVYYC-----ARGWLNMWGGTLVTV 112  
 QY 120 SS 121  
 DB 113 SS 114

Search completed: October 5, 2004, 08:01:56  
 Job time : 12.6356 secs

This Page Blank (usptg)

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd

OM protein - protein search, using sw model

Run on: October 5, 2004, 07:34:11 ; Search time 68.8117 Seconds  
(without alignments)  
596.081 Million cell updates/second

Title: US-09-805-290A-23

Sequence: 1 QVQLDSDGGGLVQAGSLRL...WGQGTQVTVSSEPKTPKPPQ 130

Scoring table: BLOSUM62

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%

Database :

```

1:  sp. archaea: *
2:  sp. bacteriela: *
3:  sp. fungi: *
4:  sp. human: *
5:  sp. invertebrate: *
6:  sp. mammal: *
7:  sp. mhc: *
8:  sp. organelle: *
9:  sp. phage: *
10: sp. plant: *
11: sp. rodent: *
12: sp. virus: *
13: sp. vertebrate: *
14: sp. unclassified: *
15: sp. virus: *
16: sp. bacteriela: *
17: sp. archaea: *

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysts of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	359	52.4	597	4	Q95B89	Q95Bb9 homo sapien
2	358.5	52.3	118	4	Q9U172	Q9U172 homo sapien
3	352	51.4	573	4	Q8WJ38	Q8WJ38 homo sapien
4	351	51.2	121	4	Q9UL71	Q9U171 homo sapien
5	349	50.9	113	4	Q9UL90	Q9U190 homo sapien
6	348.5	50.3	499	4	Q8NSK4	Q8NSK4 homo sapien
7	343	50.1	131	4	Q9UL88	Q9U188 homo sapien
8	341.5	49.6	112	4	Q9HCC1	Q9HCC1 homo sapien
9	340.5	49.7	613	4	Q8WJ1	Q8WJ1 homo sapien
10	339.5	48.5	456	4	Q95X68	Q95X68 homo sapien
11	338.3	48.5	118	4	Q9UL91	Q9U191 homo sapien
12	335	48.9	186	11	Q91Z07	Q91Z07 mus musculi
13	334.5	48.8	122	4	Q9UB44	Q9UB44 homo sapien
14	333.5	48.7	487	11	Q95XA4	Q95XA4 mus musculi
15	331	48.3	479	11	Q9U1P5	Q9U1P5 mus musculi
16	330	48.2	116	4	Q9UJ33	Q9UJ33 homo sapien

17	326.5	47.7	473	11	Q91205	Q91205	mus	musculus
18	332.2	47.0	119	11	Q920E7	Q920E7	mus	musculus
19	319.5	46.6	437	11	Q919A4	Q919A4	mus	musculus
20	319	46.6	147	1	Q950Y9	Q950Y9	homo	sapien
21	319	46.6	487	11	Q80E17	Q80E17	mus	musculus
22	318	46.4	480	11	Q91XE1	Q91XE1	mus	musculus
23	315.5	46.1	95	4	Q9ULB6	Q91XE1	mus	musculus
24	312.5	45.6	298	11	Q9GYF0	Q91XE1	mus	musculus
25	311.5	45.5	469	11	Q88YV9	Q9GYF0	mus	musculus
26	311.5	45.5	493	4	Q8NC16	Q88YV9	mus	musculus
27	307.5	44.9	521	4	Q8N4Y9	Q8NC16	homo	sapien
28	304.5	44.5	484	11	Q8V9A0	Q8N4Y9	homo	sapien
29	303.5	44.3	470	4	Q7Z5V1	Q8V9A0	homo	sapien
30	293.5	42.8	479	11	Q7TKK4	Q7Z5V1	homo	sapien
31	286	41.8	104	4	Q9UL87	Q7TKK4	mus	musculus
32	273.5	39.9	124	6	Q9N0M6	Q9UL87	homo	sapien
33	271.5	39.6	124	6	Q9N0M4	Q9N0M6	oryz	oryzolepis
34	269.5	39.3	482	4	Q7Z3S1	Q9N0M4	homo	sapien
35	267	39.0	112	4	Q9UGP3	Q7Z3S1	homo	sapien
36	266.5	38.9	241	11	Q921A6	Q9UGP3	mus	musculus
37	261.5	38.2	441	11	Q91X92	Q921A6	mus	musculus
38	253	36.9	142	11	Q924Q1	Q91X92	mus	musculus
39	251.5	36.7	124	4	Q9UL52	Q924Q1	homo	sapien
40	251.5	36.7	482	11	Q8K172	Q9UL52	mus	musculus
41	251	36.6	143	11	Q991A6	Q8K172	mus	musculus
42	249.5	36.4	143	11	Q924P9	Q991A6	mus	musculus
43	248	36.2	159	4	Q960S0	Q924P9	homo	sapien
44	247.5	36.1	145	11	Q924Q9	Q960S0	mus	musculus
45	246.5	36.0	145	11	Q924Q6	Q924Q9	mus	musculus

## ALIGNMENTS

## RESULT 1

ID	Q96BB9	PRELIMINARY;	PRT;	597	AA.
----	--------	--------------	------	-----	-----

DT 01-DEC-2001 (TREMblrel. 19, Created)  
DT 01-DEC-2001 (TREMblrel. 19, last sequence update)  
DT 01-OCT-2003 (TREMblrel. 25, last annotation update)  
D3 Hypochemical protein.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
OX NCBI\_TaxId=9606;  
RN [1]  
RP SEQUENCE FROM N.A.

SQ SEQUENCE 597 AA; 65039 MW; 4FCA3AD8ECE263D9 CRC64;

Query Match	52.4%;	Score 359;	DB 4;	Length 597;
Best Local Similarity	63.5%;	Pred. No. 4.4e-28;		
Matches	80;	Mismatches	28;	Indels 6; Gaps 3;

[illegible]

QY 116 QVTWSS 121  
 DB 139 LVTWSS 144

## RESULT 2

Q9UL72 PRELIMINARY; PRT; 118 AA.

AC Q9UL72;  
 DT 01-MAY-2000 (Tremblrel. 13, Created)  
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)  
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)  
 DE Myosin-reactive immunoglobulin heavy chain variable region  
 DE (Fragment).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 NCBI\_TaxID=9606;  
 RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=98277139; PubMed=9614934;

RA Wu X., Liu B., Van der Merwe P.L., Kalis N.N., Berner S.M.,

RT "Myosin-reactive autoantibodies in rheumatic carditis and normal

RL Clin. Immunol. Immunopathol. 87:184-192 (1998).

DR EMBL: AF035043; AAD56278.1; -.

DR PIR; S21205; S21205.

DR HSSP; P01772; 2FB4.

DR InterPro; IPR007110; IG\_1like.

DR InterPro; IPR003596; IG\_V.

DR Pfam; PF00047; IG\_1.

DR SMART; SM00406; IG\_1.

DR PROSITE; PS50835; IG\_LIKE; 1.

FT NON\_TER 1 1

FT NON\_TER 118 118

SQ SEQUENCE 118 AA; 12872 MW; B4DLA5944B2DSCA CRC64;

Query Match 52.3%; Score 358.5; DB 4; Length 118;

Best Local Similarity 64.8%; Pred. No. 5.5e-29;

Matches 79; Conservative 8; Mismatches 30; Indels 5; Gaps 3;

QY 120 SS 121

DB 117 SS 118

## RESULT 3

Q9WU38 PRELIMINARY; PRT; 573 AA.

AC Q9WU38;  
 DT 01-MAR-2002 (Tremblrel. 20, Created)  
 DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)  
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)  
 DE Hypothetical protein.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 NCBI\_TaxID=9606;  
 RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Testis;

RA Strausberg R.;

RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL; BC021276; AAH21276.1; -.  
 DR PIR; S21205; S21205.  
 DR InterPro; IPR007110; IG\_1like.  
 DR InterPro; IPR003006; IG\_MHC.  
 DR InterPro; IPR003596; IG\_V.  
 DR Pfam; PF00047; IG\_1.  
 DR SMART; SM00406; IG\_1.  
 DR PROSITE; PS50835; IG\_LIKE; 4.  
 DR PROSITE; PS00290; IG\_MHC; 2.  
 KM Hypothetical protein.  
 SQ SEQUENCE 573 AA; 62967 MW; FD072344033AC530 CRC64;

Query Match 51.4%; Score 352; DB 4; Length 573;

Best Local Similarity 59.7%; Pred. No. 2.2e-27;

Matches 80; Conservative 10; Mismatches 40; Indels 4; Gaps 3;

QY 1 QVQLDSSGGGLVQAGSLRLSCAASGTTGDIYTMAMHROAPGKERELVASAT-ESGSPNY 59

DB 1 EVQLVESGGGLVQPGSLRLSCAASGFTFDYAMHWVRQAPGKLEWVSIGISGSSIGY 79

QY 60 ADPVKGRFTISRDNKSLYLQNSLRPAEDTALYCAKKGSSGSIYIGYGMVWGQGTIV 117

DB 80 ADPVKGRFTISRDNKSLYLQNSLRPAEDTALYCAKKGSSGSIYIGYGMVWGQGTIV 139

QY 118 TVSSSEP-KTPKPOP 130

DB 140 TVSSAPTKAPDVPF 153

## RESULT 4

Q9UL71 PRELIMINARY; PRT; 121 AA.

AC Q9UL71;  
 DT 01-MAY-2000 (Tremblrel. 13, Created)  
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)  
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)  
 DE Myosin-reactive immunoglobulin heavy chain variable region  
 DE (Fragment).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 NCBI\_TaxID=9606;  
 RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=98277139; PubMed=9614934;

RA Wu X., Liu B., Van der Merwe P.L., Kalis N.N., Berner S.M.,

RT "Myosin-reactive autoantibodies in rheumatic carditis and normal

RL Clin. Immunol. Immunopathol. 87:184-192 (1998).

DR EMBL; AF035043; AAD56279.1; -.

DR HSSP; P01772; 2FB4.

DR InterPro; IPR007110; IG\_1like.

DR InterPro; IPR003596; IG\_V.

DR Pfam; PF00047; IG\_1.

DR SMART; SM00406; IG\_1.

DR PROSITE; PS50835; IG\_LIKE; 1.

FT NON\_TER 1 1

FT NON\_TER 121 121

SQ SEQUENCE 121 AA; 13154 MW; 2F045CCFASD50736 CRC64;

Query Match 51.2%; Score 351; DB 4; Length 121;

Best Local Similarity 61.3%; Pred. No. 3.4e-28;

Matches 76; Conservative 14; Mismatches 28; Indels 6; Gaps 3;

QY 1 QVQLDSSGGGLVQAGSLRLSCAASGTTGDIYTMAMHROAPGKERELVASAT-ESGSPNY 59

DB 1 EVQLVESGGGLVQPGSLRLSCAASGFTFDYAMHWVRQAPGKLEWVSIGISGSSIGY 60

QY 60 ADPVKGRFTISRDNKSLYLQNSLRPAEDTALYCAKKGSSGSIYIGYGMVWGQGTIV 117

DB 61 ADPVKGRFTISRDNKSLYLQNSLRPAEDTALYCAKKGSSGSIYIGYGMVWGQGTIV 117

```

QY      118 TVSS 121
      118 TVSS 121
DB      118 TVSS 121

RESULT 5
Q9UL90
ID Q9UL90      PRELIMINARY;      PRT;      113 AA.
AC Q9UL90;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, last annotation update)
DE Myosin-reactive immunoglobulin heavy chain variable region
DE (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OC NCBI_TaxId=9606;
[1]
RN SEQUENCE FROM N.A.
RX MEDLINE=98277139; PubMed=9614934;
RA Wu X., Liu B., Van der Merwe P.L., Kalis N.N., Berney S.M.,
RT "Myosin-reactive autoantibodies in rheumatic carditis and normal
RT fetus.";
RL Clin. Immunol. Immunopathol. 87:184-192(1998).
DR EMBL; AF035024; AAD56260.1; -.
DR HSSP; P01772; 2F84.
DR InterPro; IPR007110; IG_1like.
DR InterPro; IPR003596; IG_v.
DR Pfam; PF00047; IG; 1.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS50835; IG_LIKE; 1.
DR NON_TER 1
FT NON_TER 1
SQ SEQUENCE 113 AA; 12437 MW; ED57FPDD19086D07F CRC64;

Query Match
Best Local Similarity 50.9%; Score 349; DB 4; Length 113;
Matches 78; Conservative 6; Mismatches 28; Indels 10; Gaps 3;

QY      120 SS 121
      112 SS 113
DB      112 SS 113

RESULT 6
Q8NSK4
ID Q8NSK4      PRELIMINARY;      PRT;      499 AA.
AC Q8NSK4;
DT 01-OCT-2002 (TREMBlrel. 22, Created)
DT 01-OCT-2002 (TREMBlrel. 22, last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, last annotation update)
DE Hypothetical protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OC NCBI_TaxId=9606;
[1]
RN SEQUENCE FROM N.A.
RX MEDLINE=98277139; PubMed=9614934;
RA Wu X., Liu B., Van der Merwe P.L., Kalis N.N., Berney S.M.,
RT "Myosin-reactive autoantibodies in rheumatic carditis and normal
RT fetus.";
RL Clin. Immunol. Immunopathol. 87:184-192(1998).
DR EMBL; AF035026; AAD56262.1; -.
DR HSSP; P01810; 2F84.
DR InterPro; IPR007110; IG_1like.
DR InterPro; IPR003596; IG_v.
DR Pfam; PF00047; IG; 1.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS50835; IG_LIKE; 1.
DR NON_TER 1
FT NON_TER 1
SQ SEQUENCE 131 AA; 14142 MW; 96E7D668E375DEA0 CRC64;

Query Match
Best Local Similarity 50.1%; Score 343; DB 4; Length 131;
Matches 78; Conservative 10; Mismatches 28; Indels 20; Gaps 3;

QY      1 QVQLDSSGGGLVQAGGSLRLSCAASGTTGDIYTMAMHQAQPKERELVASAT-ESGSPVY 59
      1 EVQLVESGGGVVQPGGSLRLSCAASGFTFSSYGMHWQAQPKGLEWAFIRYDGSNKRY 60
DB      61 ADPVKGRFTISRDNGKLTIVYIQMNSLKPEDTAVVYCNALIRKFTSEYNEYWGQGTQVTV 119
      61 ADPVKGRFTISRDNGKLTIVYIQMNSLKPEDTAVVYCNALIRKFTSEYNEYWGQGTQVTV 119
QY      60 ADPVKGRFTISRDNGKLTIVYIQMNSLKPEDTAVVYCNALIRKFTSEYNEYWGQGTQVTV 119
      61 ADPVKGRFTISRDNGKLTIVYIQMNSLKPEDTAVVYCNALIRKFTSEYNEYWGQGTQVTV 119
DB      61 ADPVKGRFTISRDNGKLTIVYIQMNSLKPEDTAVVYCNALIRKFTSEYNEYWGQGTQVTV 119
      61 ADPVKGRFTISRDNGKLTIVYIQMNSLKPEDTAVVYCNALIRKFTSEYNEYWGQGTQVTV 119
QY      120 SS 121
      112 SS 113
DB      112 SS 113

RESULT 6
Q8NSK4
ID Q8NSK4      PRELIMINARY;      PRT;      499 AA.
AC Q8NSK4;
DT 01-OCT-2002 (TREMBlrel. 22, Created)
DT 01-OCT-2002 (TREMBlrel. 22, last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, last annotation update)
DE Hypothetical protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OC NCBI_TaxId=9606;
[1]
RN SEQUENCE FROM N.A.
RX MEDLINE=98277139; PubMed=9614934;
RA Wu X., Liu B., Van der Merwe P.L., Kalis N.N., Berney S.M.,
RT "Myosin-reactive autoantibodies in rheumatic carditis and normal
RT fetus.";
RL Clin. Immunol. Immunopathol. 87:184-192(1998).
DR EMBL; AF035026; AAD56262.1; -.
DR HSSP; P01810; 2F84.
DR InterPro; IPR007110; IG_1like.
DR InterPro; IPR003596; IG_v.
DR Pfam; PF00047; IG; 1.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS50835; IG_LIKE; 1.
DR NON_TER 1
FT NON_TER 1
SQ SEQUENCE 131 AA; 14142 MW; 96E7D668E375DEA0 CRC64;

Query Match
Best Local Similarity 50.1%; Score 343; DB 4; Length 131;
Matches 78; Conservative 10; Mismatches 28; Indels 20; Gaps 3;

QY      1 QVQLDSSGGGLVQAGGSLRLSCAASGTTGDIYTMAMHQAQPKERELVASAT-ESGSPVY 57
      1 EVQLVESGGGVVQPGGSLRLSCAASGFTFSSYGMHWQAQPKGLEWAFIRYDGSNKRY 60
DB      58 NVADPVKGRFTISRDNGKLTIVYIQMNSLKPEDTAVVYCNALIRKFTSEYNEYWGQGTQVTV 105
      58 NVADPVKGRFTISRDNGKLTIVYIQMNSLKPEDTAVVYCNALIRKFTSEYNEYWGQGTQVTV 105

```

```

DR      InterPro; IPR007110; IG_1like.
DR      InterPro; IPR003597; IG_CL.
DR      InterPro; IPR003006; IG_MHC.
DR      InterPro; IPR003596; IG_v.
DR      Pfam; PF00047; IG; 4.
DR      SMART; SM00409; IG; 4.
DR      SMART; SM00407; IGCL; 2.
DR      SMART; SM00406; IGV; 1.
DR      PROSITE; PS50835; IG_LIKE; 4.
DR      PROSITE; PS00290; IG_MHC; 1.
DR      Hypothetical protein.
SQ      SEQUENCE 499 AA; 53376 MW; 93AC89582054F32 CRC64;

Query Match
Best Local Similarity 50.3%; Score 344.5; DB 4; Length 499;
Matches 79; Conservative 14; Mismatches 36; Indels 9; Gaps 4;

QY      1 QVQLDSSGGGLVQAGGSLRLSCAASGTTGDIYTMAMHQAQPKERELVASAT-ESGSPVY 59
      1 EVQLVESGGGVVQPGGSLRLSCAASGFTFSSYGMHWQAQPKGLEWAFIRYDGSNKRY 79
DB      20 EVQLVESGGGVVQPGGSLRLSCAASGFTFSSYGMHWQAQPKGLEWAFIRYDGSNKRY 79
      20 EVQLVESGGGVVQPGGSLRLSCAASGFTFSSYGMHWQAQPKGLEWAFIRYDGSNKRY 79
QY      60 ADPVKGRFTISRDNGKLTIVYIQMNSLKPEDTAVVYCNALIRKFTSEYNEYWGQ 113
      60 ADPVKGRFTISRDNGKLTIVYIQMNSLKPEDTAVVYCNALIRKFTSEYNEYWGQ 113
DB      80 ADPVKGRFTISRDNGKLTIVYIQMNSLKPEDTAVVYCNALIRKFTSEYNEYWGQ 138
      80 ADPVKGRFTISRDNGKLTIVYIQMNSLKPEDTAVVYCNALIRKFTSEYNEYWGQ 138
QY      114 GTQVTVSS-EPKPKPOP 130
      139 GTVTVSSASPTSPKVPF 156
DB      139 GTVTVSSASPTSPKVPF 156

RESULT 7
Q9UL88
ID Q9UL88      PRELIMINARY;      PRT;      131 AA.
AC Q9UL88;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, last annotation update)
DE Myosin-reactive immunoglobulin heavy chain variable region
DE (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OC NCBI_TaxId=9606;
[1]
RN SEQUENCE FROM N.A.
RX MEDLINE=98277139; PubMed=9614934;
RA Wu X., Liu B., Van der Merwe P.L., Kalis N.N., Berney S.M.,
RT "Myosin-reactive autoantibodies in rheumatic carditis and normal
RT fetus.";
RL Clin. Immunol. Immunopathol. 87:184-192(1998).
DR EMBL; AF035026; AAD56262.1; -.
DR HSSP; P01810; 2F84.
DR InterPro; IPR007110; IG_1like.
DR InterPro; IPR003596; IG_v.
DR Pfam; PF00047; IG; 1.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS50835; IG_LIKE; 1.
DR NON_TER 1
FT NON_TER 1
SQ SEQUENCE 131 AA; 14142 MW; 96E7D668E375DEA0 CRC64;

Query Match
Best Local Similarity 57.2%; Score 344.5; DB 4; Length 131;
Matches 79; Conservative 14; Mismatches 36; Indels 9; Gaps 4;

QY      1 QVQLDSSGGGLVQAGGSLRLSCAASGTTGDIYTMAMHQAQPKERELVASAT-ESGSPVY 57
      1 EVQLVESGGGVVQPGGSLRLSCAASGFTFSSYGMHWQAQPKGLEWAFIRYDGSNKRY 60
DB      58 NVADPVKGRFTISRDNGKLTIVYIQMNSLKPEDTAVVYCNALIRKFTSEYNEYWGQGTQVTV 105
      58 NVADPVKGRFTISRDNGKLTIVYIQMNSLKPEDTAVVYCNALIRKFTSEYNEYWGQGTQVTV 105

```

Db 61 DYAPVKGRLTISRDSKNTLYLRNNSLKTEDTAVVYCTGTMTLVITTSKRTSF-- 118  
 QY 106 EYENYWGQGTQVTVSS 121  
 119 ---EYWGQGTLYTVSS 131

## RESULT 8

ID Q9HCC1 PRELIMINARY; PRT; 112 AA.  
 AC Q9HCC1  
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Single chain Fv (Fragment).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 NCBI\_TaxID=9606;  
 RX [1]  
 RP SEQUENCE FROM N.A.  
 RA Kikuchi M., Takeda C., Tsujimoto Y., Asada S., Nagata K.,  
 RT "An antibody fragment2A3 specific for native lysozyme: Isolation from a  
 human synthetic phage display library and characterization."  
 RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AB049915; BAB16829.1; -  
 DR HSSP: P01772; 2F84.  
 DR InterPro: IPR007110; IG-like.  
 DR InterPro: IPR003596; IG\_V.  
 DR Pfam: PF00047; IG\_1.  
 DR SMART: SM00406; IG\_1.  
 DR PROSITE: PS50835; IG\_LIKE; 1.  
 DR NON TER 1  
 FT NON TER 112  
 SO SEQUENCE 112 AA; 12243 MW; 24F1A45EC3B84788 CRC64;

Query Match 49.9%; Score 341.5; DB 4; Length 112;  
 Best Local Similarity 61.0%; Pred. No. 2.9e-27;  
 Matches 72; Conservative 13; Mismatches 26; Indels 7; Gaps 3;

QY 1 QVQLDSSGGGLVQAGSLRLSCAASGTTIDITMAMHROAPGKERELVASAT-ESGSPNY 59  
 1 EVQLVGGGGLVQPGGSLRLSCAASGFTPDYGMVWVQAPGKGLIEWVSGINMGSGTGY 60  
 60 ADPVKGRFTISRDNGLTYVYLQMSLKPEDTAVVYCNALIRKFTSEYENYWGQGTQV 117  
 61 ADSVKGRFTISRDNKNSLYLQMSLRADDTAVVYC---ARRRYAL---DYWGQGTLY 112

## RESULT 9

ID Q8WUK1 PRELIMINARY; PRT; 613 AA.  
 AC Q8WUK1  
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Hypothetical protein.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 NCBI\_TaxID=9606;  
 RX [1]  
 RP SEQUENCE FROM N.A.  
 RA Tissue-Tonsil;  
 RA Strausberg R.;  
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: BC020240; AAH20240.1; -  
 DR PIR: P10120; P10120.  
 DR PIR: S15590; S15590.  
 DR InterPro: IPR007110; IG-like.  
 DR InterPro: IPR003006; IG\_MHC.  
 DR InterPro: IPR003596; IG\_V.  
 DR Pfam: PF00047; IG\_5.

DR SMART: SM00406; IG\_V. 1.  
 DR PROSITE: PS50835; IG\_LIKE; 5.  
 DR PROSITE: PS00290; IG\_MHC; 3.  
 KW Hypothetical protein.  
 SO SEQUENCE 613 AA; 67296 MW; 60C7F5950671E315 CRC64;

## Query Match

49.7%; Score 340.5; DB 4; Length 613;  
 Best Local Similarity 64.0%; Pred. No. 3.6e-26;  
 Matches 80; Conservative 5; Mismatches 31; Indels 9; Gaps 3;

QY 1 QVQLDSSGGGLVQAGSLRLSCAASGTTIDITMAMHROAPGKERELVASAT-ESGSPNY 59  
 20 EVQLVGGGGLVQPGGSLRLSCAASGFTPDYGMVWVQAPGKGLIEWVSGINMGSGTGY 79  
 QY 60 ADPVKGRFTISRDNGLTYVYLQMSLKPEDTAVVYCNALIRKFTSEYENYWGQGTQ 116  
 80 ADSVKGRFTISRDNKNSLYLQMSLRADDTAVVYC---ADWSEGVETPDIMGQGTW 134  
 QY 117 VTWSS 121  
 135 VTWSS 139

## RESULT 10

ID Q96K68 PRELIMINARY; PRT; 494 AA.  
 AC Q96K68  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Hypothetical protein FLJ14473.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 NCBI\_TaxID=9606;  
 RX [1]  
 RP SEQUENCE FROM N.A.  
 RA Tissue-Mammary gland;  
 RA Iseogi T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,  
 RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,  
 RA Magatsuma K., Hosoi T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,  
 RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takisuchi S.,  
 RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,  
 RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahara K., Masuno Y.,  
 RA Minomiya K., Wakamatsu T.;  
 RT "NDO human cDNA sequencing project."  
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AK027379; BAB5072.1; -  
 DR PIR: S21205; S21205.  
 DR InterPro: IPR007110; IG-like.  
 DR InterPro: IPR003006; IG\_MHC.  
 DR InterPro: IPR003596; IG\_V.  
 DR Pfam: PF00047; IG\_4.  
 DR SMART: SM00406; IG\_1.  
 DR PROSITE: PS50835; IG\_LIKE; 4.  
 DR PROSITE: PS00290; IG\_MHC; 1.  
 KW Hypothetical protein.  
 SO SEQUENCE 494 AA; 53088 MW; 9A1D7AB5AEBA4C0E CRC64;

Query Match 49.5%; Score 339; DB 4; Length 494;  
 Best Local Similarity 57.7%; Pred. No. 3.9e-26;  
 Matches 79; Conservative 12; Mismatches 34; Indels 12; Gaps 4;

QY 1 QVQLDSSGGGLVQAGSLRLSCAASGTTIDITMAMHROAPGKERELVAS-ATESGSPNY 59  
 20 EVQLVGGGGLVQPGGSLRLSCAASGFTPDYGMVWVQAPGKGLIEWVSGINMGSGTGY 79  
 QY 60 ADPVKGRFTISRDNGLTYVYLQMSLKPEDTAVVYCNALIRKFTSEYENYWGQGTQ 114  
 80 ADSVKGRFTISRDNKNSLYLQMSLRADDTAVVYCNALIRKFTSEYENYWGQGTQ 134  
 QY 115 TQVTVSS-EPKTPKPOP 130

Db 135 TLVTVSASPTSPKVP 151

## RESULT 11

ID Q9UL91 PRELIMINARY; PRT; 118 AA.

AC Q9UL91;

DT 01-MAY-2000 (TREMBLrel. 13, Created)

DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)

DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)

DE Myosin-reactive immunoglobulin heavy chain variable region (Fragment).

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=98277139; PubMed=9614934;

RA Wu X., Liu B., Van der Merwe P.L., Kalis N.N., Berney S.M.,

RT "Myosin-reactive autoantibodies in rheumatic carditis and normal

RT fetus.";

RL Clin. Immunol. Immunopathol. 87:184-192(1998).

DR EMBL; AF035023; AAD56259.1; -.

DR PIR; S21205; S21205.

DR HSSP; P01772; 2FB4.

DR InterPro; IPR007110; Ig-like.

DR InterPro; IPR003596; Ig\_v.

DR Pfam; PF00047; Ig\_1.

DR SMART; SM00406; IGV; 1.

DR PROSITE; PS50835; IG\_LIKE; 1.

DR NON TER 1

FT NON TER 1

SQ SEQUENCE 118 AA; 12843 MW; D0633949P2AC149D CRC64;

Query Match 49.4%; Score 338.5; DB 4; Length 118;

Best Local Similarity 62.0%; Pred. No. 6.2e-27;

Matches 75; Conservative 12; Mismatches 29; Indels 5; Gaps 2;

QY 1 QVQLDSSGGGLVQAGSLRLSCAASGTTGDIYTMAMHROAPGKERELVASATES-GSPNY 59

Db 1 EVOQLVESGGGLVQPGSRLSLSCAASGTFSSYSMMWVQAQEGLEWVAISSTITITTY 60

QY 60 ADPVKGRFTISRDKGLTIVYLQWNSLKPEDTAVYYCNALIRKFTSEYNEYWGQGYTV 119

Db 61 ADVKGRFTISRDKGLTIVYLQWNSLKPEDTAVYYCNALIRKFTSEYNEYWGQGYTV 116

QY 120 S 120

Db 117 S 117

QY 117 S 117

Db 117 S 117

QY 117 S 117

Db 117 S 117

QY 117 S 117

Db 117 S 117

QY 117 S 117

Db 117 S 117

QY 117 S 117

Db 117 S 117

QY 117 S 117

Db 117 S 117

QY 117 S 117

Db 117 S 117

QY 117 S 117

Db 117 S 117

QY 117 S 117

Db 117 S 117

QY 117 S 117

Db 117 S 117

QY 117 S 117

Db 117 S 117

QY 117 S 117

Db 117 S 117

QY 117 S 117

Db 117 S 117

QY 117 S 117

Db 117 S 117

QY 117 S 117

Db 117 S 117

QY 117 S 117

Db 117 S 117

QY 117 S 117

Db 117 S 117

QY 117 S 117

Db 117 S 117

QY 117 S 117

Db 117 S 117

DR Pfam; PF00047; Ig; 4.

DR SMART; SM00406; IGV; 1.

DR PROSITE; PS50835; IG\_LIKE; 4.

DR PROSITE; PS00290; IG\_MHC; 2.

KW Hypothetical protein.

SQ SEQUENCE 486 AA; 52682 MW; 4FEF835125DA870B CRC64;

Query Match 48.9%; Score 335; DB 11; Length 486;

Best Local Similarity 53.5%; Pred. No. 9.8e-26;

Matches 68; Conservative 18; Mismatches 33; Indels 8; Gaps 2;

QY 1 QVQLDSSGGGLVQAGSLRLSCAASGTTGDIYTMAMHROAPGKERELVASATESGSPNY 60

Db 20 EYLVESGGGLVQPGSRLSLSCAASGTFSSYSMMWVQAQEGLEWVAISSTITITTY 79

QY 61 DPKGRFTISRDKGLTIVYLQWNSLKPEDTAVYYCNALIRKFTSEYNEYWGQGYTV 115

Db 80 DNVKGRFTISRDKGLTIVYLQWNSLKPEDTAVYYCNALIRKFTSEYNEYWGQGYTV 136

QY 116 QVTSSSE 122

Db 137 TIVTSSSE 143

QY 137 TIVTSSSE 143

Db 137 TIVTSSSE 143

QY 137 TIVTSSSE 143

Db 137 TIVTSSSE 143

QY 137 TIVTSSSE 143

Db 137 TIVTSSSE 143

QY 137 TIVTSSSE 143

Db 137 TIVTSSSE 143

QY 137 TIVTSSSE 143

Db 137 TIVTSSSE 143

QY 137 TIVTSSSE 143

Db 137 TIVTSSSE 143

QY 137 TIVTSSSE 143

Db 137 TIVTSSSE 143

QY 137 TIVTSSSE 143

Db 137 TIVTSSSE 143

QY 137 TIVTSSSE 143

Db 137 TIVTSSSE 143

QY 137 TIVTSSSE 143

Db 137 TIVTSSSE 143

QY 137 TIVTSSSE 143

Db 137 TIVTSSSE 143

QY 137 TIVTSSSE 143

Db 137 TIVTSSSE 143

QY 137 TIVTSSSE 143

Db 137 TIVTSSSE 143

QY 137 TIVTSSSE 143

Db 137 TIVTSSSE 143

QY 137 TIVTSSSE 143

Db 137 TIVTSSSE 143

QY 137 TIVTSSSE 143

Db 137 TIVTSSSE 143

QY 137 TIVTSSSE 143

Db 137 TIVTSSSE 143

QY 137 TIVTSSSE 143

Db 137 TIVTSSSE 143

QY 137 TIVTSSSE 143

Db 137 TIVTSSSE 143

QY 137 TIVTSSSE 143

Db 137 TIVTSSSE 143

QY 137 TIVTSSSE 143

Db 137 TIVTSSSE 143

QY 137 TIVTSSSE 143

Db 137 TIVTSSSE 143

QY 137 TIVTSSSE 143

Db 137 TIVTSSSE 143

QY 137 TIVTSSSE 143

Db 137 TIVTSSSE 143

QY 137 TIVTSSSE 143

Db 137 TIVTSSSE 143

QY 137 TIVTSSSE 143

Db 137 TIVTSSSE 143

QY 137 TIVTSSSE 143

Db 137 TIVTSSSE 143

QY 137 TIVTSSSE 143

Db 137 TIVTSSSE 143

QY 137 TIVTSSSE 143

Db 137 TIVTSSSE 143

QY 137 TIVTSSSE 143

Db 137 TIVTSSSE 143

QY 137 TIVTSSSE 143

Db 137 TIVTSSSE 143

QY 137 TIVTSSSE 143

Db 137 TIVTSSSE 143

QY 137 TIVTSSSE 143

Db 137 TIVTSSSE 143

QY 137 TIVTSSSE 143

Db 137 TIVTSSSE 143

```

RESULT 14
Q99XA4      PRELIMINARY;      FRT;      487 AA.
ID Q99XA4;
AC Q99XA4;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC004786; AA04786.1; -.
DR HSSP; P01810; 2FBJ.
DR InterPro; IPR007110; Ig_Like.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR003596; Ig_V.
DR Pfam; PF00047; Ig; 4.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS50835; IG_LIKE; 4.
DR PROSITE; PS00290; IG_MHC; 2.
KM Hypothetical protein.
SQ SEQUENCE 487 AA; 52554 MW; 7DC9E96DB33077B CRC64;

Query Match      48.7%; Score 333.5; DB 11; Length 487;
Best Local Similarity 57.0%; Pred. No. 1.4e-25;
Matches 73; Conservative 13; Mismatches 33; Indels 9; Gaps 3;

QY 1 OVOLODSGGIVQAGSLRLSCAASGTLGDIYTMAMHROAPGKERELVASATESGS-PNY 59
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 20 EVQLVESGGGLVAPGSLKVSCHASGFTFSSTAMSWROTPKRLRWVAITSDGSSITYY 79
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 60 ADPVKGRFTISRDNGLTVYLQNSLKPEDTAVYYCNALIRKFTSEYNEVWGQGTQVTV 114
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 80 PDVTKGRFTISRDNAGNLYLQMSHLKSRDTAVYYC--ARDMGSPYGYSRFDYWGQG 136
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 115 TQTVSSSE 122
   :|||:|||||
Db 137 TTVVSSSE 144
   :|||:|||||

RESULT 15
Q91WPS      PRELIMINARY;      FRT;      479 AA.
ID Q91WPS;
AC Q91WPS;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC013656; AA013656.1; -.
DR InterPro; IPR007110; Ig_Like.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR003596; Ig_V.
DR Pfam; PF00047; Ig; 4.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS50835; IG_LIKE; 4.
DR PROSITE; PS00290; IG_MHC; 2.
KM Hypothetical protein.
SQ SEQUENCE 479 AA; 51603 MW; ECB2D0877748584F CRC64;

```

```

Query Match      48.3%; Score 331; DB 11; Length 479;
Best Local Similarity 55.6%; Pred. No. 2.5e-25;
Matches 69; Conservative 16; Mismatches 31; Indels 8; Gaps 2;

QY 1 OVOLODSGGIVQAGSLRLSCAASGTLGDIYTMAMHROAPGKERELVASA-TEGSSPNY 59
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 20 EVQLVESGGGLVAPGSLKVSCHASGFTFSSTAMSWROTPKRLRWVAITSDGSSITYY 79
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 60 ADPVKGRFTISRDNGLTVYLQNSLKPEDTAVYYCNALIRKFTSEYNEVWGQGTQVTV 119
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 80 SDTWKGRFTISRDNAGKSTLYLQMSLSRSEDYAVYYC-----VRGGYFDVWGAGTAVTV 132
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 120 SSSE 123
   :|||
Db 133 SSSE 136
   :|||

```

Search completed: October 5, 2004, 08:13:47  
 Job time : 69.8117 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OW protein - protein search, using sw model

Run on: October 5, 2004, 07:04:10 ; Search time 98.9043 Seconds  
(without alignments)  
371.381 Million cell updates/sec

Title: US-09-805-290A-23

Perfect score: 685  
Sequence: 1 QVQLQDSGGGLVQAGGSLRL.....WGQGTQVTVSEPKPKPP 130

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Genesep28Jan04:\*  
1: Genesep28Jan04:\*  
2: Genesep28Jan04:\*  
3: Genesep28Jan04:\*  
4: Genesep28Jan04:\*  
5: Genesep28Jan04:\*  
6: Genesep28Jan04:\*  
7: Genesep28Jan04:\*  
8: Genesep28Jan04:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	685	100.0	130	4	AAE10556 HPL inhib
2	563	82.2	130	4	AAE10557 HPL inhib
3	522	76.2	130	4	AAE10552 HPL inhib
4	518.5	75.7	129	4	AAE10553 HPL inhib
5	513	74.9	130	4	AAE10554 HPL inhib
6	502.5	73.4	131	4	AAE10555 HPL inhib
7	500	73.0	130	4	AAE10559 HPL inhib
8	494	72.1	124	4	AAE10563 HPL inhib
9	491.5	71.8	129	4	AAE10558 HPL inhib
10	491	71.7	124	4	AAE10566 HPL inhib
11	487	71.1	152	4	AAE105283 HPL inhib
12	482	70.4	130	4	AAE10567 HPL inhib
13	473.5	69.1	131	5	ABG30620 Immunoglo
14	463.5	67.7	129	4	AAE10561 HPL inhib
15	459.5	67.1	129	4	AAE10564 HPL inhib
16	459.5	67.1	149	4	AAE10568 HPL inhib
17	454.5	66.4	129	4	AAE10551 HPL inhib
18	451.5	65.9	133	5	ABG30872 Immunoglo
19	451	65.8	128	4	AAE10565 HPL inhib
20	443.5	64.7	377	4	AAE105286 HPL inhib
21	443.5	64.7	383	4	AAE105287 HPL inhib
22	443	64.7	190	7	ABR62883 Llama ant
23	442	64.3	190	7	ABR62882 Llama ant
24	440.5	64.3	124	2	AAE10565 HPL inhib
25	440	64.2	134	5	ABG30618 Immunoglo

26	440	64.2	190	7	ABR62881 Llama ant
27	437	63.8	190	7	ABR62879 Llama ant
28	436.5	63.7	225	2	AAE10565 HPL inhib
29	436	63.6	190	7	ABR62880 Llama ant
30	435.5	63.6	219	2	AAE10565 HPL inhib
31	434	63.4	194	7	ABR62885 Llama ant
32	432	62.9	138	5	ABG30619 Immunoglo
33	431	62.9	198	7	ABR62878 Llama ant
34	430	62.8	127	4	AAE10562 HPL inhib
35	429.5	62.7	127	5	ABG30617 Mouse cro
36	429	62.6	124	4	AAE10562 HPL inhib
37	428	62.5	119	4	AAE10562 HPL inhib
38	427.5	62.4	204	2	AAE10562 HPL inhib
39	425	62.0	119	4	AAE10562 HPL inhib
40	423.5	61.8	118	2	AAE10562 HPL inhib
41	423.5	61.8	193	7	ABR62884 Llama ant
42	422	61.8	120	4	AAE10560 HPL inhib
43	422.5	61.7	205	2	AAE10562 HPL inhib
44	419.5	61.2	500	3	AAE10562 HPL inhib
45	419.5	61.2	500	4	AAE10562 HPL inhib

## ALIGNMENTS

RESULT 1	AAE10556	AAE10556 standard; peptide; 130 AA.
ID	AAE10556	AAE10556 standard; peptide; 130 AA.
AC	AAE10556	AAE10556 standard; peptide; 130 AA.
DT	10-DEC-2001	(first entry)
DE	HPL inhibiting VHH fragment, HPL #18 from llama species.	
XX	Llama antibody; camelid; anorectic; heavy chain variable domain; VHH;	
KM	human dietary enzyme inhibitor; medicament; human pancreatic lipase; HPL;	
KM	food; human gastric lipase; HGL; cosmetic control; body weight.	
XX	Llama sp.	
OS	Llama sp.	
XX	Key	Location/Qualifiers
FT	Region	31..35
FT	Region	/label=CDR1
FT	Region	/note="Complementarity determining region 1"
FT	Region	/label=CDR2
FT	Region	/note="Complementarity determining region 2"
FT	Region	/label=CDR3
FT	Region	/note="Complementarity determining region 3"
XX	Ep1134231-A1.	
XX	19-SEP-2001.	
XX	20-FEB-2001; 2001EP-00200703.	
XX	14-MAR-2000; 2000EP-00200930.	
XX	(UNIT) UNILEVER NV.	
XX	(UNIT) UNILEVER P.L.C.	
XX	Beemer S, Van De Burg M, De Haard JWM, Tareilus B;	
XX	WPI; 2001-572718/65.	
XX	New antibody or its fragments for inhibiting human dietary enzymes,	
XX	useful for cosmetic control of body weight of human beings, comprises	
XX	heavy chain variable domain derived from immunoglobulin naturally devoid	
XX	of light chains.	
XX	Example 2, Page 10; 37pp; English.	

XX The patent discloses antibodies or their fragments comprising a heavy  
 CC chain variable domain (VH) derived from an immunoglobulin naturally  
 CC devoid of light chains specific for inhibiting human dietary enzymes. The  
 CC antibodies of the invention are useful for the preparation of medicaments  
 CC or food for inhibiting the activity of one or more human dietary enzymes  
 CC especially human pancreatic lipase (HPL) or human gastric lipase (HGL)  
 CC which are useful for the cosmetic control of body weight of human beings.  
 CC The present peptide sequence is HPL inhibiting VH fragment, HPL #18 from  
 CC llama (camelid) species  
 XX  
 SQ Sequence 130 AA;  
 Query Match 100.0%; Score 695; DB 4; Length 130;  
 Best Local Similarity 100.0%; Pred. No. 2.3e-53;  
 Matches 130; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 QVQLDSGGGLVQAGSLRLSCAASGTDITVMAMHQAQPKERELVASATESGSPNYA 60  
 DB 1 QVQLDSGGGLVQAGSLRLSCAASGTDITVMAMHQAQPKERELVASATESGSPNYA 60  
 QY 61 DPVKGFTISRDNKGLTVYLQWNSLKPEDTAVYYCNALIRKFTSEYNEVWGQGTQVTVS 120  
 DB 61 DPVKGFTISRDNKGLTVYLQWNSLKPEDTAVYYCNALIRKFTSEYNEVWGQGTQVTVS 120  
 QY 121 SEPKTPKPOP 130  
 DB 121 SEPKTPKPOP 130  
 RESULT 2  
 AAE10557  
 ID AAE10557 standard; peptide, 130 AA.  
 AC AAE10557;  
 XX  
 DT 10-DEC-2001 (first entry)  
 XX  
 DE HPL inhibiting VH fragment, HPL #19 from llama species.  
 XX  
 KW llama antibody; camelid; anorectic; heavy chain variable domain; VH;  
 KW human dietary enzyme inhibitor; medicament; human pancreatic lipase; HPL;  
 KW food; human gastric lipase; HGL; cosmetic control; body weight.  
 XX  
 OS llama sp.  
 XX  
 FH Key Location/Qualifiers  
 FT Region 31..35  
 FT /label= CDR1  
 FT /note= "Complementarity determining region 1"  
 FT Region 50..64  
 FT /label= CDR2  
 FT /note= "Complementarity determining region 2"  
 FT Region 98..110  
 FT /label= CDR3  
 FT /note= "Complementarity determining region 3"  
 XX  
 PN EP1134231-A1.  
 XX  
 PD 19-SEP-2001.  
 XX  
 PF 20-FEB-2001; 2001EP-00200703.  
 XX  
 PR 14-MAR-2000; 2000EP-00200930.  
 XX  
 PA (UNITL ) UNILEVER NV.  
 PA (UNITL ) UNILEVER PLC.  
 XX  
 PI Bezemer S, Van De Burg M, De Haard JWM, Tareilus E;  
 XX  
 DR WPI; 2001-572718/65.  
 XX  
 RT New antibody or its fragments for inhibiting human dietary enzymes,

PT useful for cosmetic control of body weight of human beings, comprises  
 PT heavy chain variable domain derived from immunoglobulin naturally devoid  
 PT of light chains.  
 XX  
 PS Example 2; Page 10; 37pp; English.  
 XX  
 CC The patent discloses antibodies or their fragments comprising a heavy  
 CC chain variable domain (VH) derived from an immunoglobulin naturally  
 CC devoid of light chains specific for inhibiting human dietary enzymes. The  
 CC antibodies of the invention are useful for the preparation of medicaments  
 CC or food for inhibiting the activity of one or more human dietary enzymes  
 CC especially human pancreatic lipase (HPL) or human gastric lipase (HGL)  
 CC which are useful for the cosmetic control of body weight of human beings.  
 CC The present peptide sequence is HPL inhibiting VH fragment, HPL #19 from  
 CC llama (camelid) species  
 XX  
 SQ Sequence 130 AA;  
 Query Match 82.2%; Score 563; DB 4; Length 130;  
 Best Local Similarity 80.0%; Pred. No. 1.7e-42;  
 Matches 104; Conservative 10; Mismatches 16; Indels 0; Gaps 0;  
 QY 1 QVQLDSGGGLVQAGSLRLSCAASGTDITVMAMHQAQPKERELVASATESGSPNYA 60  
 DB 1 QVQLDSGGGLVQAGSLRLSCAASGTDITVMAMHQAQPKERELVASATESGSPNYA 60  
 QY 61 DPVKGFTISRDNKGLTVYLQWNSLKPEDTAVYYCNALIRKFTSEYNEVWGQGTQVTVS 120  
 DB 61 DPVKGFTISRDNKGLTVYLQWNSLKPEDTAVYYCNALIRKFTSEYNEVWGQGTQVTVS 120  
 QY 121 SEPKTPKPOP 130  
 DB 121 SEPKTPKPOP 130  
 RESULT 3  
 AAE10552  
 ID AAE10552 standard; peptide, 130 AA.  
 AC AAE10552;  
 XX  
 DT 10-DEC-2001 (first entry)  
 XX  
 DE HPL inhibiting VH fragment, HPL #12 from llama species.  
 XX  
 KW llama antibody; camelid; anorectic; heavy chain variable domain; VH;  
 KW human dietary enzyme inhibitor; medicament; human pancreatic lipase; HPL;  
 KW food; human gastric lipase; HGL; cosmetic control; body weight.  
 XX  
 OS llama sp.  
 XX  
 FH Key Location/Qualifiers  
 FT Region 31..35  
 FT /label= CDR1  
 FT /note= "Complementarity determining region 1"  
 FT Region 50..64  
 FT /label= CDR2  
 FT /note= "Complementarity determining region 2"  
 FT Region 98..110  
 FT /label= CDR3  
 FT /note= "Complementarity determining region 3"  
 XX  
 PN EP1134231-A1.  
 XX  
 PD 19-SEP-2001.  
 XX  
 PF 20-FEB-2001; 2001EP-00200703.  
 XX  
 PR 14-MAR-2000; 2000EP-00200930.  
 XX  
 PA (UNITL ) UNILEVER NV.  
 PA (UNITL ) UNILEVER PLC.  
 XX

Pt	Bezemer S, Van De Burg M, De Haard JWM, Tareilus F;
Dd	WPI; 2001-572718/65.
Pt	New antibody or its fragments for inhibiting human dietary enzymes,
Pt	useful for cosmetic control of body weight of human beings, comprises
Pt	heavy chain variable domain derived from immunoglobulin naturally devoid
Pt	of light chains.
Ps	Example 2; Page 9; 37pp; English.
Cc	The patent discloses antibodies or their fragments comprising a heavy
Cc	chain variable domain (VH) derived from an immunoglobulin naturally
Cc	devoid of light chains specific for inhibiting human dietary enzymes. The
Cc	antibodies of the invention are useful for the preparation of medicaments
Cc	or food for inhibiting the activity of one or more human dietary enzymes
Cc	especially human pancreatic lipase (HPL) or human gastric lipase (HGL)
Cc	which are useful for the cosmetic control of body weight of human beings.
Cc	The present peptide sequence is HPL inhibiting VH fragment, HPL #12 from
Cc	llama (camelid) species
Sq	Sequence 130 AA;
Query Match	76.2%; Score 522; DB 4; Length 130;
Best Local Similarity	77.7%; Pred. No. 7.9e-39;
Matches 101; Conservative	9; Mismatches 20; Indels 0; Gaps 0;
Qy	1 QVOLODSGGGVQAGGSLRLCAASGTIGDIYTMAMHQAPPKERELVASATSGSPNYA 60
Db	1 QVOLODSGGGVQAGGSURLCSAHSGSIGSIHTMGWVYQTPKERDVAITLDDGSTIYA 60
Qy	61 DPAKGRFTISSDNGLIVYLQNNLSIKPEDTAVYVCNALIRKFTSEYNWGQTGYTS 120
Db	61 DSVKGRFTISRNDNTINTVYLQNNDLKPEDTAVYCNADVPRPTRSYLEVMQGTLVTYS 120
Qy	121 SEPTKPPROP 130 
Db	121 SEPTKPPROP 130 
RESULT 4	
AAB10553	ID AAE10553 standard; peptide; 129 AA.
AAE10553;	
10-DEC-2001	(first entry)
HPL inhibiting VHH fragment, HPL #13 from llama species.	
Llama sp.	
Location/Qualifiers	
/label=.35	
/note="Complementarity determining region 1"	
/label=.64	
/label=CDR2	
/note="Complementarity determining region 2"	
/label=.109	
/note="Complementarity determining region 3"	
EPI134231-A1.	
19-SEP-2001.	
20-FEB-2001; 2001EP-00200703.	

PR	14-MAR-2000; 2000EP-00200930.
XX	(UNITL ) UNILEVER NV.
PA	(UNITL) UNILEVER PLC.
XX	
P1	Bezemer S, Van De Burg M, De Haard JW, Tarelius E;
XX	
DR	WPI; 2001-572718/65.
PT	New antibody or its fragments for inhibiting human dietary enzymes,
PT	useful for cosmetic control of body weight of human beings, compris
PT	heavy chain variable domain derived from immunoglobulin naturally devoid
PT	of light chains.
XX	
PS	Example 2; Page 9; 37pp; English.
XX	
CC	The patent discloses antibodies or their fragments comprising a heavy
CC	chain variable domain (VH) derived from an immunoglobulin naturally
CC	devoid of light chains specific for inhibiting human dietary enzymes. The
CC	antibodies of the invention are useful for the preparation of medicaments
CC	or food for inhibiting the activity of one or more human dietary enzymes
CC	especially human pancreatic lipase (HPL) or human gastric lipase (HGL)
CC	which are useful for the cosmetic control of body weight of human beings.
CC	The present peptide sequence is HPL inhibiting VH fragment, HPL #13 from
CC	Llama (camelid) species
SQ	Sequence 129 AA;
	Query Match                75.7%; Score 518.5; DB 4; Length 129;
	Best Local Similarity     78.5%; Pred. No. 1.6e-38;
	Matches 102; Conservative   6; Mismatches 21; Indels 1; Gaps 1
Qy	1 QVQLDPSGGSLVQAAGSRLSCAASGTTIGDIYTWAMRQAPGKERELVASATESGSPNYA 60
Dd	1 QVQLSGSSGGGVIAQAGSRLSCAASGTILSIIVMDWRQIPGKRELNVRITAGSTNYA 60
Qy	61 DPKRGFTISRNKGKLTYLLQNSLKRPEDNAVYYCNALLRKRTSENEYWGCGTGYTYS 120
Dd	61 DSAKGFITISKDNANRYLVLMNSLKPEDNAVYYCNALLTR-WDSKYNDVWGCGTGYTYS 119
Qy	121 SEPKTPKPOP 130
Dd	120 SEPKTPKPOP 129
RESULT 5	
AAE10554	
ID	AAE10554 standard; peptide; 110 AA.
AC	AAE10554;
XX	
DT	10-DEC-2001 (first entry)
XX	
DE	HPL inhibiting VH fragment, HPL #14 from llama species.
KW	Llama antibody; camelid; anorectic; heavy chain variable domain; VH;
KW	human dietary enzyme inhibitor; medicament; human pancreatic lipase; HPL;
KW	food; human gastric lipase; HGL; cosmetic control; body weight.
OS	Lama sp.
XX	
Key	Location/Qualifiers
FT	Region                31..35
FT	/label= CDR1
FT	/note= "Complementarity determining region 1"
FT	Region                50..64
FT	/label= CDR2
FT	/note= "Complementarity determining region 2"
FT	Region                98..110
FT	/label= CDR3
FT	/note= "Complementarity determining region 3"
BN	EPI134231-AI.

```

XX 19-SEP-2001.
PD 20-FEB-2001; 2001EP-00200703.
XX PF
XX 14-MAR-2000; 2000EP-00200930.
XX PR
XX (UNIL ) UNILEVER NV.
XX PA (UNIL ) UNILEVER PLC.
XX PI
XX Bezemer S, Van De Burg M, De Haard JWW, Tareilus E;
XX WPI; 2001-572718/65.
XX DR
XX New antibody or its fragments for inhibiting human dietary enzymes,
XX PT useful for cosmetic control of body weight of human beings, comprises
XX PT heavy chain variable domain derived from immunoglobulin naturally devoid
XX PT of light chains.
XX PS
XX Example 2; Page 9; 37pp; English.
XX CC The patent discloses antibodies or their fragments comprising a heavy
XX CC chain variable domain (VH) derived from an immunoglobulin naturally
XX CC devoid of light chains specific for inhibiting human dietary enzymes. The
XX CC antibodies of the invention are useful for the preparation of medicaments
XX CC or food for inhibiting the activity of one or more human dietary enzymes
XX CC especially human pancreatic lipase (HPL) or human gastric lipase (HGL)
XX CC which are useful for the cosmetic control of body weight of human beings.
XX CC The present peptide sequence is HPL inhibiting VH fragment, HPL #14 from
XX CC llama (camelid) species
XX SQ Sequence 130 AA;

Query Match 74.9%; Score 513; DB 4; Length 130;
Best Local Similarity 76.3%; Pred. No. 5e-38; Indels 0; Gaps 0;
Matches 100; Conservative 10; Mismatches 20;

QY 1 QVQLQDSGGGLVQAGSLRLSCAASGTDIDITMAMHROAPGKERELVASATESGSPNYA 60
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 QVQLQDSGGGLVQAGSLRLSCAASGSISSITMGWRFQPKQRELVASITSGSTNYA 60
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 61 DPKVGRFTISRDNKLTLYLQNNSLKPEPTAVYYCNALIRKFTSEY-NEWYQGTQVTV 120
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 61 DSVKGRFTISRDNILNTVYLQNNSLKPEPTAVYHCNADVPRYTSRYELWQGTLTWVS 120
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 121 SEPKTPKPOP 130
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 121 SEPKTPKPOP 130
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 6
AAE10555
ID AAE10555 standard; peptide; 131 AA.
XX
AC AAE10555;
XX
DT 10-DEC-2001 (first entry)
XX
DE HPL inhibiting VH fragment, HPL #15 from llama species.
XX
KW llama antibody; camelid; anorectic; heavy chain variable domain; VH;
KW human dietary enzyme inhibitor; medicament; human pancreatic lipase; HPL;
KW food; human gastric lipase; HGL; cosmetic control; body weight.
XX
OS Lama sp.
XX
FH Key Location/Qualifiers
FT 31..35
FT /label= CDR1
FT /note= "Complementarity determining region 1"
FT Region 50..64
FT /label= CDR2
FT /note= "Complementarity determining region 2"

```

```

FT Region 98..111
FT /label= CDR3
FT /note= "Complementarity determining region 3"
XX
XX EPI134231-A1.
XX
XX 19-SEP-2001.
XX PD
XX 20-FEB-2001; 2001EP-00200703.
XX PF
XX 14-MAR-2000; 2000EP-00200930.
XX PR
XX (UNIL ) UNILEVER NV.
XX PA (UNIL ) UNILEVER PLC.
XX PI
XX Bezemer S, Van De Burg M, De Haard JWW, Tareilus E;
XX WPI; 2001-572718/65.
XX DR
XX New antibody or its fragments for inhibiting human dietary enzymes,
XX PT useful for cosmetic control of body weight of human beings, comprises
XX PT heavy chain variable domain derived from immunoglobulin naturally devoid
XX PT of light chains.
XX PS
XX Example 2; Page 9; 37pp; English.
XX CC The patent discloses antibodies or their fragments comprising a heavy
XX CC chain variable domain (VH) derived from an immunoglobulin naturally
XX CC devoid of light chains specific for inhibiting human dietary enzymes. The
XX CC antibodies of the invention are useful for the preparation of medicaments
XX CC or food for inhibiting the activity of one or more human dietary enzymes
XX CC especially human pancreatic lipase (HPL) or human gastric lipase (HGL)
XX CC which are useful for the cosmetic control of body weight of human beings.
XX CC The present peptide sequence is HPL inhibiting VH fragment, HPL #15 from
XX CC llama (camelid) species
XX SQ Sequence 131 AA;

Query Match 73.4%; Score 502.5; DB 4; Length 131;
Best Local Similarity 76.3%; Pred. No. 4.3e-37; Indels 1; Gaps 1;
Matches 100; Conservative 7; Mismatches 23;

QY 1 QVQLQDSGGGLVQAGSLRLSCAASGTDIDITMAMHROAPGKERELVASATESGSPNYA 60
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 QVQLQDSGGGLVQAGSLRLSCAASGISINVMGFQPKQRELVASITSGSTNYA 60
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 61 DPKVGRFTISRDNKLTLYLQNNSLKPEPTAVYYCNALIRKFTSEY-NEWYQGTQVTV 119
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 61 DSLKGRFTISRDNKAAVAVYLNQNNLKPEDTAVYYCNALITPAGSSNYVYGVWGHGTXVTV 120
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 120 SEPKTPKPOP 130
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 121 SEPKTPKPOP 131
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 7
AAE10559
ID AAE10559 standard; peptide; 130 AA.
XX
AC AAE10559;
XX
DT 10-DEC-2001 (first entry)
XX
DE HPL inhibiting VH fragment, HPL #30 from llama species.
XX
KW llama antibody; camelid; anorectic; heavy chain variable domain; VH;
KW human dietary enzyme inhibitor; medicament; human pancreatic lipase; HPL;
KW food; human gastric lipase; HGL; cosmetic control; body weight.
XX
OS Lama sp.
XX
FH Key Location/Qualifiers
FT 31..35
FT Region

```

```

FT FT /label= CDR1 /note= "Complementarity determining region 1"
FT FT Region 49.64
FT FT /label= CDR2
FT FT /note= "Complementarity determining region 2"
FT FT Region 98.110
FT FT /label= CDR3
FT FT /note= "Complementarity determining region 3"
FN FN EPI134231-A1.
XX XX 19-SEP-2001.
PD PD 20-FEB-2001; 2001EP-00200703.
PF PF 14-MAR-2000; 2000EP-00200930.
PR PR
XX XX (UNITL ) UNILEVER NV.
XX XX (UNITL ) UNILEVER PLC.
PI PI Bezemer S, Van De Burg M, De Haard JMW, Tarelus E,
XX XX WPI; 2001-572718/65.
XX XX
XX XX
XX XX New antibody or its fragments for inhibiting human dietary enzymes,
PT PT useful for cosmetic control of body weight of human beings, comprises
PT PT heavy chain variable domain derived from immunoglobulin naturally devoid
PT PT of light chains.
PS PS
PS PS Example 2; Page 10; 37pp; English.
XX XX
XX XX The patent discloses antibodies or their fragments comprising a heavy
CC CC chain variable domain (VH) derived from an immunoglobulin naturally
CC CC devoid of light chains specific for inhibiting human dietary enzymes. The
CC CC antibodies of the invention are useful for the preparation of medicaments
CC CC or food for inhibiting the activity of one or more human dietary enzymes
CC CC especially human pancreatic lipase (HPL) or human gastric lipase (HGL)
CC CC which are useful for the cosmetic control of body weight of human beings.
CC CC The present peptide sequence is HPL inhibiting VH fragment, HPL #30 from
XX XX llama (camelid) species
XX XX
SQ SQ Sequence 130 AA;
Query Match 73.0%; Score 500; DB 4; Length 130;
Best Local Similarity 75.4%; Pred. No. 7.2e-37;
Matches 98; Conservative 8; Mismatches 24; Indels 0; Gaps 0
QY 1 QVQLQDSGGGLVOAGSGLRSLCAASGTTIGDLYTMMARHQAQGRKRELVASATSSGSPNTA 60
DB 1 QVQLQDSGGGLVOAGSGLRSLCAASGSLGDIHTMGWRQTPGRKRDVVAITQSGSSTNTYA 60
QY 61 DPVKGRFTTISRDNKLTVTYLQNNLSIKPEDTAVVYCNALIRKFTSEYNEWGGGTQVTVS 120
DB 61 DSVKGRFTTISRDNLTNTLVYLQNNDLKPEDTGYYWNADVPRYSRLFIWGGTLVTF 120
QY 121 SEPKTEKPPQ 130
DB 121 LEPKTRKPPQ 130
XX XX
XX XX HGL inhibiting VH fragment, HGL #9 from llama species.
DE DE
XX XX
XX XX Llama antibody; camelid; anorectic; heavy chain variable domain; VH;
XX XX human dietary enzyme inhibitor; medicament; human pancreatic lipase; HPL;
XX XX food; human gastric lipase; HGL; cosmetic control; body weight.

```

XX	Lama sp.
OS	
XX	
XX	Location/Qualifiers
FR	Key
FT	Region
FT	31..35
FT	/label= CDR1
FT	/note= "Complementarity determining region 1"
FT	50..64
FT	/label= CDR2
FT	/note= "Complementarity determining region 2"
FT	98..104
FT	/label= CDR3
FT	/note= "Complementarity determining region 3"
PN	
XX	EP1134231-AI.
XX	
XX	19-SEP-2001.
PF	
XX	20-FEB-2001; 2001EP-00200703.
PR	
XX	14-MAR-2000; 2000EP-00200930.
PA	(UNITL ) UNILEVER NV.
XX	(UNITL ) UNILEVER PLC.
PI	
XX	Bezemer S, Van De Burg M, De Haard JW, Tarelius E;
DR	WPI; 2001-572718/65.
PT	
PT	New antibody or its fragments for inhibiting human dietary enzymes,
PT	useful for cosmetic control of body weight of human beings, comprises
PT	heavy chain variable domain derived from immunoglobulin naturally devoid
XX	of light chains.
PS	
XX	Example 4; Page 13; 37pp: English.
CC	The patent discloses antibodies or their fragments comprising a heavy
CC	chain variable domain (VH) derived from an immunoglobulin naturally
CC	devoid of light chains specific for inhibiting human dietary enzymes. The
CC	antibodies of the invention are useful for the preparation of medicaments
CC	or food for inhibiting the activity of one or more human dietary enzymes
CC	especially human pancreatic lipase (HPL) or human gastric lipase (HGL)
CC	which are useful for the cosmetic control of body weight of human beings.
CC	The present peptide sequence is HGL inhibiting VH fragment, HGL #9 from
CC	llama (camelid) species
XX	
SQ	Sequence 124 AA;
	Query Match            72.1%;   Score 494;   DB 4;   Length 124;
	Best Local Similarity   74.6%;   Pred. No. 2.3e-36;
	Matches   97;   Conservative   8;   Mismatches   19;   Indels   6;   Gaps   1
OY	1 QVOLQDSGGGLVOAGSGSLRLSCAASGTIGDIYTAMWRQAPGKERELVASATSGSPNYA 60
Dd	1 QVLQDSGGGLVQAQGSRLSCAASGSIQSYVMWRQAPGRREVAALMGSSGTYYA 60
OY	61 DPKRGFTISRNGKLTLYLLQNNISKPEDTAVTYCNALIRRKFTSEYNEYWGCGITVTVS 120
Dd	61 DSKRGFTISRDNINRTMYLLQNNISKPEDTGVVYCAG-----TGABGHYWGGGTIVTS 114
OY	121 SEPKTPKPQP 130
Dd	115 SEPKTPKPQP 124
RESULT 9	
AAE10558	
ID AAE10558	standard, peptide; 129 AA.
XX	AAE10558;
XX	
XT	10-DEC-2001 (first entry)
XX	

DE HPL inhibiting VHH fragment, HPL #22 from llama species.  
 XX  
 XX Llama antibody; camelid; anorectic; heavy chain variable domain; VHH;  
 KW human dietary enzyme inhibitor; medicament; human pancreatic lipase; HPL;  
 KW food; human gastric lipase; HGL; cosmetic control; body weight.  
 XX  
 OS Lama sp.  
 XX  
 XX Key Location/Qualifiers  
 FH 31..35  
 FT Region /label= CDR1  
 FT /note= "Complementarity determining region 1"  
 FT 50..64  
 FT Region /label= CDR2  
 FT /note= "Complementarity determining region 2"  
 FT 98..109  
 FT Region /label= CDR3  
 FT /note= "Complementarity determining region 3"  
 XX  
 XX EPI134231-A1.  
 PN 19-SEP-2001.  
 PD 20-FEB-2001; 2001EP-00200703.  
 PF 14-MAR-2000; 2000EP-00200930.  
 PR (UNIL ) UNILEVER NV.  
 PA (UNIL ) UNILEVER PLC.  
 PI Bezemer S, Van De Burg M, De Haard JWW, Tareilus E;  
 PI WPI; 2001-572718/65.  
 DR  
 XX New antibody or its fragments for inhibiting human dietary enzymes,  
 PT useful for cosmetic control of body weight of human beings, comprises  
 PT heavy chain variable domain derived from immunoglobulin naturally devoid  
 PT of light chains.  
 XX  
 XX Example 2; Page 10; 37pp; English.  
 PS  
 XX The patent discloses antibodies or their fragments comprising a heavy  
 CC chain variable domain (VHH) derived from an immunoglobulin naturally  
 CC devoid of light chains specific for inhibiting human dietary enzymes. The  
 CC antibodies of the invention are useful for the preparation of medicaments  
 CC or food for inhibiting the activity of one or more human dietary enzymes  
 CC especially human pancreatic lipase (HPL) or human gastric lipase (HGL)  
 CC which are useful for the cosmetic control of body weight of human beings.  
 CC The present peptide sequence is HPL inhibiting VHH fragment, HPL #22 from  
 CC llama (camelid) species  
 CC  
 CC Sequence 129 AA;  
 SQ  
 Query Match 71.8%; Score 491.5; DB 4; Length 129;  
 Best Local Similarity 73.8%; Pred. No. 4.1e-36;  
 Matches 96; Conservative 11; Mismatches 22; Indels 1; Gaps 1;  
 XX  
 QY 1 QVQLDSSGGGLVQAGSLRLSCAASGTLIDITMAMHROAPGKERELVASATSSGSPNYA 60  
 DB 1 QVQLDSSGGGLVQAGSLRLSCAASGTLISIMTWYRQAPGERELVARMSSDGTTSY 60  
 QY 61 DPKVGRFTISRDNKGLTYVLQNNSLKPEDTAVYYCNALLRRKFTSEYNEYWGCGTQYTVS 120  
 DB 61 DSKVGRFTISRDNKGLTYVLMNNLKPEDTAVYYCALL-SSYDGSWMDYGGGQGTQYTVS 119  
 QY 121 SEPTPKPOP 130  
 DB 120 SEPTPKPOP 129  
 RESULT 10  
 AAE10566  
 ID AAE10566 standard; peptide; 124 AA.

XX AAE10566;  
 AC  
 XX 10-DEC-2001 (first entry)  
 DT  
 XX HGL inhibiting VHH fragment, HGL #15 from llama species.  
 DE  
 XX Llama antibody; camelid; anorectic; heavy chain variable domain; VHH;  
 KW human dietary enzyme inhibitor; medicament; human pancreatic lipase; HPL;  
 KW food; human gastric lipase; HGL; cosmetic control; body weight.  
 XX  
 OS Lama sp.  
 XX  
 XX Key Location/Qualifiers  
 FH 30..35  
 FT Region /label= CDR1  
 FT /note= "Complementarity determining region 1"  
 FT 50..64  
 FT Region /label= CDR2  
 FT /note= "Complementarity determining region 2"  
 FT 98..104  
 FT Region /label= CDR3  
 FT /note= "Complementarity determining region 3"  
 XX  
 XX EPI134231-A1.  
 PN 19-SEP-2001.  
 PD 20-FEB-2001; 2001EP-00200703.  
 PF 14-MAR-2000; 2000EP-00200930.  
 PR (UNIL ) UNILEVER NV.  
 PA (UNIL ) UNILEVER PLC.  
 PI Bezemer S, Van De Burg M, De Haard JWW, Tareilus E;  
 PI WPI; 2001-572718/65.  
 DR  
 XX New antibody or its fragments for inhibiting human dietary enzymes,  
 PT useful for cosmetic control of body weight of human beings, comprises  
 PT heavy chain variable domain derived from immunoglobulin naturally devoid  
 PT of light chains.  
 XX  
 XX Example 4; Page 14; 37pp; English.  
 PS  
 XX The patent discloses antibodies or their fragments comprising a heavy  
 CC chain variable domain (VHH) derived from an immunoglobulin naturally  
 CC devoid of light chains specific for inhibiting human dietary enzymes. The  
 CC antibodies of the invention are useful for the preparation of medicaments  
 CC or food for inhibiting the activity of one or more human dietary enzymes  
 CC especially human pancreatic lipase (HPL) or human gastric lipase (HGL)  
 CC which are useful for the cosmetic control of body weight of human beings.  
 CC The present peptide sequence is HGL inhibiting VHH fragment, HGL #15 from  
 CC llama (camelid) species  
 CC  
 CC Sequence 124 AA;  
 SQ  
 Query Match 71.7%; Score 491, DB 4; Length 124;  
 Best Local Similarity 73.8%; Pred. No. 4.3e-36;  
 Matches 96; Conservative 8; Mismatches 20; Indels 6; Gaps 1;  
 XX  
 QY 1 QVQLDSSGGGLVQAGSLRLSCAASGTLIDITMAMHROAPGKERELVASATSSGSPNYA 60  
 DB 1 QVQLDSSGGGLVQAGSLRLSCAASGTLISIMTWYRQAPGERELVARMSSDGTTSY 60  
 QY 61 DPKVGRFTISRDNKGLTYVLQNNSLKPEDTAVYYCNALLRRKFTSEYNEYWGCGTQYTVS 120  
 DB 61 DSKVGRFTISRDNKGLTYVLMNNLKPEDTAVYYCALL-SSYDGSWMDYGGGQGTQYTVS 114  
 QY 121 SEPTPKPOP 130  
 DB 115 SEPTPKPOP 124

RESULT 11  
AAE05283  
ID AAE05283 standard; protein; 152 AA.  
XX  
AC AAE05283;  
XX  
DT 18-SEP-2001 (first entry)  
XX  
DE Anti-potato SBEII (Clone46) VH region attached with myc and his6 tag.  
XX  
KW Potato; heavy chain immunoglobulin; pathogen resistance;  
XX metabolism modulator; passive immunisation; heavy chain variable domain;  
XX VH; anti-potato SBEII; starch branching enzyme; SBE A.  
XX  
OS Solanum tuberosum.  
OS Unidentified.  
OS Chimeric.  
XX  
PN EP118669-A2.  
XX  
PD 25-JUL-2001.  
XX  
PF 08-DEC-2000; 2000EP-00310997.  
XX  
PR 17-DEC-1999; 99EP-00310188.  
XX  
PA (UNITL ) UNILEVER PLC.  
PA (UNITL ) UNILEVER NV.  
XX  
PI Frenken LGJ, Van Der Logt CPE, Jobling SA, Teh Y;  
XX  
DR WPI: 2001-427157/46.  
DR N-PSDB; AAD10054.  
XX  
PT Modifying a plant to produce an antibody useful for increasing pathogen  
PT resistance or to modulate metabolism comprises introducing a DNA sequence  
PT encoding a heavy chain immunoglobulin linked to a peptide that targets a  
PT cellular compartment.  
XX  
PS Example 1; Fig 12; 81pp; English.  
XX  
CC The present invention relates to a method for modifying a plant to  
CC produce an antibody or an active fragment or derivative, or a protein  
CC functional equivalent, in a cellular compartment comprising introducing a  
CC DNA sequence encoding a heavy chain immunoglobulin, where the DNA is  
CC linked to promoters and provided with an additional sequence encoding a  
CC peptide capable of targeting heavy chain immunoglobulin to a cellular  
CC compartment. The method is used for producing a heavy chain  
CC immunoglobulin or an active fragment or derivative, or a protein that is  
CC functionally equivalent for increasing the pathogen resistance in a plant  
CC or to modulate metabolism in a plant. Under some circumstances it may be  
CC desirable to retain the antibody product with the plant rather than  
CC extracting and isolating the product. In particular, edible selected  
CC antigen may be used in a method of passively immunising an animal. The  
CC preferably human, against the antigen, e.g., pathogenic organisms. The  
CC present sequence is an anti-potato SBEII antibody (denoted Clone46) heavy  
CC chain variable domain (VH) attached to peptide linkers, myc and his6 tag.  
CC The potato SBEII is a starch branching enzyme also designated SBE A  
CC  
SQ Sequence 152 AA;

Query Match 71.1%; Score 487; DB 4; Length 152;  
Best Local Similarity 73.7%; Pred. No. 1.2e-35;  
Matches 98; Conservative 7; Mismatches 18; Indels 10; Gaps 2;

QY 1 QVQLDSGGGLVOAGSLRLSCAASGTTIDITMAMHRAQPKERELVASATSSGSPNYA 60  
DB 1 QVQLDSGGGLVOAGSLRLSCAASGTTISITAMMYRQAPKQKEVVASINSISTNYA 60  
QY 61 DSVKGRFTISRDNKLTIVYLQMSLKPEDTAVYYC--NALIRKFTSEYNEWGGTQV 117

DB 61 DSVKGRFTISRDNKLTIVYLQMSLKPEDTAVYYCAAGNLVLR-----PYWGGTIV 113  
QY 118 TVSSSEPTKPKPP 130  
DB 114 TVSSSEPTKPKPP 126

RESULT 12  
AAE10567  
ID AAE10567 standard; peptide; 130 AA.  
XX  
AC AAE10567;  
XX  
DT 10-DEC-2001 (first entry)  
XX  
DE HGL inhibiting VH fragment, HGL #16 from llama species.  
XX  
KW Llama antibody; camelid; anorectic; heavy chain variable domain; VH;  
KW human dietary enzyme inhibitor; medicament; human pancreatic lipase; HPL;  
KW food; human gastric lipase; HGL; cosmetic control; body weight.  
XX  
OS Lama sp.  
XX  
FH Key Location/Qualifiers  
FH 31..35  
FT Region /label= CDR1  
FT /note= "Complementarity determining region 1"  
FT 50..64  
FT Region /label= CDR2  
FT /note= "Complementarity determining region 2"  
FT 98..110  
FT Region /label= CDR3  
FT /note= "Complementarity determining region 3"  
XX  
PN EP134231-A1.  
XX  
PD 19-SEP-2001.  
XX  
PF 20-FEB-2001; 2001EP-00200703.  
XX  
PR 14-MAR-2000; 2000EP-00200930.  
XX  
PA (UNITL ) UNILEVER NV.  
PA (UNITL ) UNILEVER PLC.  
XX  
PI Bezemer S, Van De Burg M, De Haard JWM, Tareilus E;  
XX  
DR WPI: 2001-572718/65.  
XX  
PT New antibody or its fragments for inhibiting human dietary enzymes,  
PT useful for cosmetic control of body weight of human beings, comprises  
PT heavy chain variable domain derived from immunoglobulin naturally devoid  
PT of light chains.  
XX  
PS Example 4; Page 14; 37pp; English.  
XX  
CC The patent discloses antibodies or their fragments comprising a heavy  
CC chain variable domain (VH) derived from an immunoglobulin naturally  
CC devoid of light chains specific for inhibiting human dietary enzymes. The  
CC antibodies of the invention are useful for the preparation of medicaments  
CC or food for inhibiting the activity of one or more human dietary enzymes  
CC especially human pancreatic lipase (HPL) or human gastric lipase (HGL),  
CC which are useful for the cosmetic control of body weight of human beings.  
CC The present peptide sequence is HGL inhibiting VH fragment, HGL #16 from  
CC llama (camelid) species  
CC  
SQ Sequence 130 AA;

Query Match 70.4%; Score 482; DB 4; Length 130;  
Best Local Similarity 73.8%; Pred. No. 2.9e-35;  
Matches 96; Conservative 10; Mismatches 24; Indels 0; Gaps 0;

QY 1 QVQLDSGGGLVOAGSLRLSCAASGTTIDITMAMHRAQPKERELVASATSSGSPNYA 60

```

DB 1 QVLOESGGGLVQAGSLRLSCAASGSDPRYNAMAYRQAPKORQLVATITYRTNYA 60
QY 61 DPVKGFTTSDNGKLTLYVLQNNSLKPEDPAVYVCNALRRKFTSYNEVWGQGTQVYV 120
DB 61 DSVKGRFTISRDNKNTLYLQNNSLKPEDPAVYVCNALRRKFTSYNEVWGQGTQVYV 120
QY 121 SSEKTPKPPQ 130
DB 121 SSEKTPKPPQ 130

RESULT 13
ABG30620 standard; protein; 131 AA.
XX ABG30620;
AC ABG30620;
XX
XX 21-OCT-2002 (first entry)
DE Immunoglobulin G specific heavy chain variable domain antibody #3.
XX
XX Heavy chain variable domain; antibody; protein array; aging; VHH;
KM Immunoglobulin; sticky phage-antibody; micro-panning; immune library;
XX proteomic; mouse.
OS Mus sp.
XX WO200248193-A2.
XX 20-JUN-2002.
XX
XX 03-DEC-2001; 2001WO-EP014471.
XX
XX 13-DEC-2000; 2000EP-00311142.
XX
XX (UNIL ) UNILEVER NV.
PA (UNIL ) UNILEVER PLC.
PA (UNIL ) HINDUSTAN LEVER LTD.
XX
XX De Haard JTW, Hermans P, Landa I, Verrips CT;
XX
XX WPI; 2002-583487/62.
XX
XX Novel protein array useful for detecting the presence of individual
PT proteins in sample, comprises heavy-chain variable domain antibodies or
PT antibody fragments obtainable from Camelidae.
XX
XX Disclosure; Fig 3; 80pp; English.
XX
XX The invention describes a protein array (I) comprising a number of heavy-
XX chain variable domain antibodies or antibody fragments, obtainable from
XX Camelidae. The method is useful for removing abundant proteins from an
XX extract or sample which do not provide useful information on the
XX condition of a cell or tissue in the extract or sample to be
XX investigated. (I) is useful for detecting the presence of individual
XX proteins in a sample, comparing the distribution of proteins in different
XX cell types, and identification of proteins that may be of importance in
XX determining the altered properties of cells in disease, aging or other
XX conditions. Using a heavy-chain variable domain derived from an
XX immunoglobulin that is naturally devoid of light chains (VHH) in (I)
XX provides a number of advantages, such as an improvement of
XX sensitivity/resolution in the order of 10-100 times, and detection of
XX post-translationally modified proteins. The invention also describes a
XX method (II) that enables the simultaneous processing of large numbers of
XX target antigens in a controlled way. The incorporated phase-ELISA
XX generates on-line information about the success or failure of a certain
XX panning condition. This feature combined with the microtiter plate format
XX allows the complete automation of the technology, based on computer-made
XX decisions on the values of the phage-ELISA for continuation of a limited
XX number of selections. In (II), due to the fact that many different
XX conditions can be tested, varying amounts of input-phages can be used
XX simultaneously in order to decrease the enrichment of sticky phage-

```

```

CC antibodies. Micro-panning is an effective tool for selecting both naive,
CC synthetic and immune libraries on large numbers of different target
CC molecules, enabling the generation of large panels of antibodies in short
CC time frames needed for the generation of arrays (proteomics). This
CC sequence represents an immunoglobulin (Ig) heavy chain variable domain
CC antibody (VHH)
XX
XX Sequence 131 AA;
SQ
Query Match 69.1%; Score 473.5; DB 5; Length 131;
Best Local Similarity 72.5%; Pred. No. 1,7e-34;
Matches 95; Conservative 6; Mismatches 29; Indels 1; Gaps 1;

QY 1 QVLOESGGGLVQAGSLRLSCAASGSDPRYNAMAYRQAPKORQLVATITYRTNYA 60
DB 1 QVLOESGGGLVQAGSLRLSCAASGSDPRYNAMAYRQAPKORQLVATITYRTNYA 60
QY 61 DPVKGFTTSDNGKLTLYVLQNNSLKPEDPAVYVCNALRRKFTSYNEVWGQGTQVYV 119
DB 61 DSVKGRFTISRDNKNTLYLQNNSLKPEDPAVYVCNALRRKFTSYNEVWGQGTQVYV 120
QY 120 SSEKTPKPPQ 130
DB 121 SSEKTPKPPQ 131

RESULT 14
AAE10561
ID AAE10561 standard; peptide; 129 AA.
XX
XX AAE10561;
AC AAE10561;
XX
XX 10-DEC-2001 (first entry)
DE HGL inhibiting VHH fragment, HGL #4 from llama species.
XX
XX llama antibody; camelid; anorectic; heavy chain variable domain; VHH;
XX human dietary enzyme inhibitor; medicament; human pancreatic lipase; HPL;
XX food; human gastric lipase; HGL; cosmetic control; body weight.
XX
XX Lama sp.
OS
XX
XX Key Location/Qualifiers
FH 31..35 FT
FT Region /label= CDR1
FT /note= "Complementarity determining region 1"
FT Region 50..64
FT /label= CDR2
FT /note= "Complementarity determining region 2"
FT 98..109
FT /label= CDR3
FT /note= "Complementarity determining region 3"
XX
XX EPI134231-A1.
XX
XX 19-SEP-2001.
XX
XX 20-FEB-2001; 2001EP-00200703.
XX
XX 14-MAR-2000; 2000EP-00200930.
XX
XX (UNIL ) UNILEVER NV.
PA (UNIL ) UNILEVER PLC.
XX
XX Bezemer S, Van De Burg M, De Haard JTW, Tareilus E;
XX
XX WPI; 2001-572718/65.
XX
XX New antibody or its fragments for inhibiting human dietary enzymes,
XX useful for cosmetic control of body weight of human beings, comprises
XX heavy chain variable domain derived from immunoglobulin naturally devoid
XX of light chains.

```



PS Example 4; Page 13; 37pp; English.

CC The patent discloses antibodies or their fragments comprising a heavy  
CC chain variable domain (VH) derived from an immunoglobulin naturally  
CC devoid of light chains specific for inhibiting human dietary enzymes. The  
CC antibodies of the invention are useful for the preparation of medicaments  
CC or food for inhibiting the activity of one or more human dietary enzymes  
CC especially human pancreatic lipase (HPL) or human gastric lipase (HGL)  
CC which are useful for the cosmetic control of body weight of human beings.  
CC The present peptide sequence is HGL inhibiting VH fragment, HGL #4 from  
CC llama (camelid) species

SO Sequence 129 AA;

Query Match 67.7%; Score 463.5; DB 4; Length 129;

Best Local Similarity 69.7%; Pred. No. 1.3e-33;

Matches 92; Conservative 12; Mismatches 23; Indels 5; Gaps 2;

```
QY 1 QVQLDSSGGGLVQAGSLRLSCAASGTTIDYTMAMHQAQPKERELVASATESGSPNYA 60
DB 1 QVQLDSSGGGLVQAGSLRLSCAASGTTIDYTMAMHQAQPKERELVASATESGSPNYA 60
QY 61 DPKGRFTISRDNKGLTYVQLQWNSLKPEDTAVYYCNALIRKFTSEYNE--YWGQGTQV 118
DB 61 DSVKGRFTIARENAKNTVYLQWNSLKPEDTAVYYCKG--RGLTQYSEHDYWGQGTQV 117
QY 119 VSSEPKTKPKQP 130
DB 118 VSSEPKTKPKQP 129
```

RESULT 15

AAE10564

ID AAE10564 standard; peptide; 129 AA.

AAE10564;

DT 10-DEC-2001 (first entry)

DE HGL inhibiting VH fragment, HGL #10 from llama species.

KW llama antibody; camelid; anorectic; heavy chain variable domain; VH;  
KW human dietary enzyme inhibitor; medicament; human pancreatic lipase; HPL;  
KW food; human gastric lipase; HGL; cosmetic control; body weight.

OS Lama sp.

Key Location/Qualifiers

FT 31..35

FT /label=CDR1

FT /note="Complementarity determining region 1"

FT 50..64

FT /label=CDR2

FT /note="Complementarity determining region 2"

FT 98..109

FT /label=CDR3

FT /note="Complementarity determining region 3"

EP134231-A1.

19-SEP-2001.

20-FEB-2001; 2001EP-00200703.

14-MAR-2000; 2000EP-00200930.

PA (UNITL) UNILEVER NV.

PA (UNITL) UNILEVER PLC.

PI Bezemer S, Van De Burg M, De Haard JW, Tareilus E;

WPI; 2001-572718/65.

PT New antibody or its fragments for inhibiting human dietary enzymes,  
PT useful for cosmetic control of body weight of human beings, comprises  
PT heavy chain variable domain derived from immunoglobulin naturally devoid  
PT of light chains.

PS Example 4; Page 13; 37pp; English.

CC The patent discloses antibodies or their fragments comprising a heavy  
CC chain variable domain (VH) derived from an immunoglobulin naturally  
CC devoid of light chains specific for inhibiting human dietary enzymes. The  
CC antibodies of the invention are useful for the preparation of medicaments  
CC or food for inhibiting the activity of one or more human dietary enzymes  
CC especially human pancreatic lipase (HPL) or human gastric lipase (HGL)  
CC which are useful for the cosmetic control of body weight of human beings.  
CC The present peptide sequence is HGL inhibiting VH fragment, HGL #10 from  
CC llama (camelid) species

SO Sequence 129 AA;

Query Match 67.1%; Score 459.5; DB 4; Length 129;

Best Local Similarity 69.7%; Pred. No. 2.9e-33;

Matches 92; Conservative 11; Mismatches 24; Indels 5; Gaps 2;

```
QY 1 QVQLDSSGGGLVQAGSLRLSCAASGTTIDYTMAMHQAQPKERELVASATESGSPNTA 60
DB 1 QVQLDSSGGGLVQAGSLRLSCAASGTTIDYTMAMHQAQPKERELVASATESGSPNTA 60
QY 61 DPKGRFTISRDNKGLTYVQLQWNSLKPEDTAVYYCNALIRKFTSEYNE--YWGQGTQV 118
DB 61 NSVKGRTISRDNKGLTYVQLQWNSLKPEDTAVYYCKG--RGLTQYSEHDYWGQGTQV 117
QY 119 VSSEPKTKPKQP 130
DB 118 VSSEPKTKPKQP 129
```

Search completed: October 5, 2004, 08:00:12

Job time: 99.5043 secs

This Page Blank (usps)

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2004, 07:36:16 ; Search time 21.0648 Seconds

(without alignments)  
593.639 Million cell updates/sec

Title: US-09-805-290A-24

Perfect score: 695  
Sequence: 1 QVQLDSGGGLVQTGSLRL.....WGQGTQVTVSSSEPKTPKPP 130

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

1: Pir1:\*  
2: Pir2:\*  
3: Pir3:\*  
4: Pir4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	378.5	54.5	120	2	S48798 Ig heavy chain V r
2	378	54.4	123	2	S31114 Ig heavy chain - h
3	373	53.7	121	2	I55673 Ig heavy chain - h
4	372	53.5	119	2	S31107 Ig heavy chain - h
5	371	53.4	119	2	S31108 Ig heavy chain - h
6	369	53.1	140	2	S31686 Ig heavy chain V r
7	369	53.1	160	2	S05271 Ig heavy chain pre
8	367	52.8	140	2	S31588 Ig heavy chain V r
9	364.5	52.4	114	2	S46391 Ig heavy chain V r
10	364.5	52.4	124	2	S20782 Ig heavy chain V r
11	363	52.2	119	2	C36005 Ig heavy chain V r
12	363	52.2	138	2	S31666 Ig heavy chain V r
13	362	52.1	119	2	D36005 Ig heavy chain V r
14	362	52.1	119	2	F36005 Ig heavy chain V r
15	361.5	52.0	120	2	S31112 Ig heavy chain - h
16	359.5	51.7	118	2	S31105 Ig heavy chain (su
17	359.5	51.7	140	2	S70442 Ig heavy chain pre
18	359	51.7	135	2	S31598 Ig heavy chain V r
19	358	51.5	125	2	S30531 Ig heavy chain V r
20	357.5	51.4	114	2	S31120 Ig heavy chain - h
21	357.5	51.4	114	2	S46390 Ig heavy chain V r
22	357	51.4	120	2	S36273 Ig heavy chain V r
23	357	51.4	139	2	I37781 Ig variable region
24	356	51.2	121	2	S19666 Ig heavy chain V r
25	356	51.2	127	2	S38489 Ig heavy chain - h
26	355	51.1	117	2	S78486 Ig heavy chain V r
27	355	51.1	121	2	S26798 Ig heavy chain V r
28	355	51.1	134	2	S31679 Ig heavy chain V r
29	354.5	51.0	114	2	S46392 Ig heavy chain V r

30	354	50.9	120	2	S55539 Ig heavy chain V r
31	354	50.9	121	2	G36005 Ig heavy chain V r
32	353.5	50.9	112	2	PH1647 Ig heavy chain V r
33	353.5	50.9	122	2	E36005 Ig heavy chain V r
34	353.5	50.9	128	2	S26790 Ig heavy chain V r
35	353.5	50.9	126	2	S31587 Ig heavy chain V r
36	353	50.8	120	2	S36278 Ig heavy chain V r
37	353	50.8	134	2	S31699 Ig heavy chain V r
38	353	50.8	140	2	A30532 Ig heavy chain pre
39	352	50.6	120	2	S55536 Ig heavy chain V r
40	352	50.6	123	2	S39493 Ig heavy chain - h
41	351.5	50.6	122	2	S20772 Ig heavy chain V r
42	351.5	50.6	147	2	I37780 Ig variable region
43	351	50.5	120	2	S55538 Ig heavy chain V r
44	351	50.5	122	2	S31603 Ig heavy chain V r
45	349.5	50.3	128	2	S26786 Ig heavy chain V r

#### ALIGNMENTS

##### RESULT 1

S48798

Ig heavy chain V region (anti-Sm, VH3/Dxp4/JH4d) - human

C/Species: Homo sapiens (man)

C/Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 23-Jul-1999

C/Accession: S48798

R/Mahmoudi, M.; Edwards, J.; Cairns, E.; Bell, D.

submitted to the EMBL Data Library, October 1994

A/Description: Molecular characterization of natural human anti-Sm autoantibodies.

A/Reference number: S48797

A/Accession: S48798

A/Status: preliminary

A/Molecule type: mRNA

A/Residues: 1-120 <MAH>

A/Cross-references: EMBL:Z46382; NID:G562324; PIDN:CA86521.1; PID:G1340167

C/Suprafamily: immunoglobulin V region; immunoglobulin homology

C/Keywords: heterotrimer; immunoglobulin

F:15-98/Domain: immunoglobulin homology <IMM>

Query Match 54.5%; Score 378.5; DB 2; Length 120;

Best Local Similarity 67.2%; Pred. No. 1.9e-25;

Matches 82; Conservative 10; Mismatches 27; Indels 3; Gaps 3;

QY 1 QVQLDSGGGLVQTGSLRLCAASGPIDGVYIMGWYRQAPGKQRMVASITGPEN-Y 59

##### RESULT 2

S31114

Ig heavy chain - human

C/Species: Homo sapiens (man)

C/Date: 02-Dec-1993 #sequence\_revision 26-May-1995 #text\_change 17-Mar-1999

C/Accession: S31114

R/Rapchort, F.M.; Timmers, E.; Kenter, M.T.H.; Van Tol, M.D.; Vossen, J.M.; Schuurman,

A/Title: Restricted utilization of germ-line V(H)3 genes and short diverse third comple

A/Reference number: S31104; M01D:9211633; PMID:1730252

A/Accession: S31114

A/Status: preliminary; nucleic acid sequence not shown; translation not shown

A/Molecule type: mRNA

A/Residues: 1-123 <RAA>

A/Cross-references: EMBL:X62963

A/Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1991

C:Superfamily: immunoglobulin V region; immunoglobulin homology  
C:Keywords: heterotetramer; immunoglobulin  
F:15-98/Domain: immunoglobulin homology <IMM>

Query Match 54.4%; Score 378; DB 2; Length 123;  
Best Local Similarity 65.0%; Pred. No. 2.1e-25;  
Matches 80; Conservative 11; Mismatches 30; Indels 2; Gaps 2;

QY 1 QVQLDSSGGGLVQPGGSLRLSCAASGPIGDIVLMGWYRQAPGKQREMVASITATGPPN-Y 59  
Db 1 EVQLLESQGGGLVQPGGSLRLSCAASGFTTSSVAMSWRQAPGKGLWVSALISGSGSTYY 60

QY 60 TDSVKGFTISRNDKNTLEYLQNNLKPEPTAVYYCAQVRFSSDYNTNWGGQTQTV 118  
Db 61 ADSVKGFTISRNDKNTLEYLQNNSLRAEDTAVYYCAKASLYLRFLEWFDYWGQGLTVT 120

QY 119 VSS 121  
Db 121 VSS 123

## RESULT 3

Ig heavy chain - human (fragment)

C:Species: Homo sapiens (man)

C:Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 21-Jan-2000

C:Accession: I55673

R:Knights, G.B.; Agnello, V.; Bonagura, V.; Barnes, J.L.; Panka, D.J.; Zhang, Q.X.

J. Exp. Med. 178, 1903-1911, 1993

A:Title: Human rheumatoid factor cross-idiotypes. IV. Studies on WA Xid-positive Igm with

clonal from the 17.109 and G6 Xids

A:Reference number: I55673; PMID:94065558; PMID:8245772

A:Accession: I55673

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-121 <RES>

A:Cross-references: GB:I587268; NID:9186197; PIDN:AAC97536.1; PID:9186198

C:Gene: GDB:I58M

A:Cross-references: GDB:I20086; OMIM:147020

A:Map position: 14q32.33-14q32.33

C:Superfamily: immunoglobulin V region; immunoglobulin homology

F:15-98/Domain: immunoglobulin homology <IMM>

Query Match 53.7%; Score 373; DB 2; Length 121;  
Best Local Similarity 65.6%; Pred. No. 5.6e-25;  
Matches 80; Conservative 9; Mismatches 31; Indels 2; Gaps 2;

QY 1 QVQLDSSGGGLVQPGGSLRLSCAASGPIGDIVLMGWYRQAPGKQREMVASITATGPPN-Y 59  
Db 1 EVQLLESQGGGLVQPGGSLRLSCAASGFTTSSVAMSWRQAPGKGLWVSALISGSGSTYY 60

QY 60 TDSVKGFTISRNDKNTLEYLQNNLKPEPTAVYYCAQVRFSSDYNTNWGGQTQTV 119  
Db 61 ADSVKGFTISRNDKNTLEYLQNNSLRAEDTAVYYCAAPRAAGSPY-DYWGQGLTVT 119

QY 120 SS 121  
Db 120 SS 121

## RESULT 4

Ig heavy chain - human

C:Species: Homo sapiens (man)

C:Date: 02-Dec-1993 #sequence\_revision 26-May-1995 #text\_change 17-Mar-1999

C:Accession: S31107

R:Rapoport, F.M.; Timmers, E.; Kenter, M.J.H.; van Tol, M.J.D.; Vossen, J.M.; Schuurman

Eur. J. Immunol. 22, 247-251, 1992

A:Title: Restricted utilization of germ-line V(H)3 genes and short diverse third comple

A:Reference number: S31104; PMID:92111633; PMID:1130252

A:Accession: S31107

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: RNA  
A:Residues: 1-119 <RAA>  
A:Cross-references: EMBL:X62955  
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1991  
C:Superfamily: immunoglobulin V region; immunoglobulin homology  
C:Keywords: heterotetramer; immunoglobulin  
F:15-98/Domain: immunoglobulin homology <IMM>

Query Match 53.5%; Score 372; DB 2; Length 119;  
Best Local Similarity 64.8%; Pred. No. 6.7e-25;  
Matches 79; Conservative 11; Mismatches 28; Indels 4; Gaps 2;

QY 1 QVQLDSSGGGLVQPGGSLRLSCAASGPIGDIVLMGWYRQAPGKQREMVASITATGPPN-Y 59  
Db 1 EVQLLESQGGGLVQPGGSLRLSCAASGFTTSSVAMSWRQAPGKGLWVSALISGSGSTYY 60

QY 60 TDSVKGFTISRNDKNTLEYLQNNLKPEPTAVYYCAQVRFSSDYNTNWGGQTQTV 119  
Db 61 ADSVKGFTISRNDKNTLEYLQNNSLRAEDTAVYYC---AKDPGASYFDYWGQGLTVT 117

QY 120 SS 121  
Db 118 SS 119

## RESULT 5

Ig heavy chain - human

C:Species: Homo sapiens (man)

C:Date: 02-Dec-1993 #sequence\_revision 26-May-1995 #text\_change 17-Mar-1999

C:Accession: S31108

R:Rapoport, F.M.; Timmers, E.; Kenter, M.J.H.; van Tol, M.J.D.; Vossen, J.M.; Schuurman,

Eur. J. Immunol. 22, 247-251, 1992

A:Title: Restricted utilization of germ-line V(H)3 genes and short diverse third comple

A:Reference number: S31104; PMID:92111633; PMID:1130252

A:Accession: S31108

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: mRNA

A:Residues: 1-119 <RAA>

A:Cross-references: EMBL:X62956

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1991

C:Superfamily: immunoglobulin V region; immunoglobulin homology

C:Keywords: heterotetramer; immunoglobulin  
F:15-98/Domain: immunoglobulin homology <IMM>

Query Match 53.4%; Score 371; DB 2; Length 119;  
Best Local Similarity 64.8%; Pred. No. 8.2e-25;  
Matches 79; Conservative 10; Mismatches 29; Indels 4; Gaps 2;

QY 1 QVQLDSSGGGLVQPGGSLRLSCAASGPIGDIVLMGWYRQAPGKQREMVASITATGPPN-Y 59  
Db 1 EVQLLESQGGGLVQPGGSLRLSCAASGFTTSSVAMSWRQAPGKGLWVSALISGSGSTYY 60

QY 60 TDSVKGFTISRNDKNTLEYLQNNLKPEPTAVYYCAQVRFSSDYNTNWGGQTQTV 119  
Db 61 ADSVKGFTISRNDKNTLEYLQNNSLRAEDTAVYYCAKDRRLTGTD---YWGQGLTVT 117

QY 120 SS 121  
Db 118 SS 119

## RESULT 6

Ig heavy chain V region - human (fragment)

C:Species: Homo sapiens (man)

C:Date: 22-Nov-1993 #sequence\_revision 10-Nov-1995 #text\_change 23-Jul-1999

C:Accession: S31686

R:Cuisinier, A.M.; Gauthier, L.; Boulli, L.; Fougereau, M.; Tonnelie, C.

submitted to the EMBL Data Library, June 1992

A:Description: Mechanisms that generate human immunoglobulin diversity operate from the

A:Reference number: S31585

A:Accession: S31686

Tue Oct 5 09:51:21 2004

us-09-805-290a-24.rpr

**Page 3**

A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-140 <CUI>  
A:Cross-references: EMBL:Z14205; NID:g30969; PIDD:CAA74574.1; PID:g30970  
C:Superfamily: immunoglobulin V region; immunoglobulin homology  
C:Keywords: heterotetramer; immunoglobulin  
F:34-117/Domain: Immunoglobulin homology <IMM>

Query Match	53.1%;	Score 369;	DB 2;	Length 140;
Best Local Similarity	64.8%;	Pred. No. 1.4e-24;		
Matches	79;	Conservative	11;	Mismatches 30;
			Indels	2;
			Gaps	2;

QY 1 QVQLDSDSGGGLVQTQGGSLRLSCAAGSPGIDGYIMQMYCAPEKQKQEMVSAIATAPPN-Y 59  
Db 20 EVQLLESQGLVQPGGSLRLSCAAGSPFFSSYAMSVNQAPKGLQEMVSAISGGSGSTYY 79  
QY 60 TDSVYGRFTISPDNKTETELONMLKPEDTAVYVCNCAQVRAFRSSDTYNWGQSTQVY 119  
Db 80 SDSVYGRFTISDNKNTLYLQMSLRADDTAVTC-AACPAGGSPFSDTIVGGGLVY 138

QY	120	SS	121
Db	139	SS	140

RESULT 7  
S05271

C/Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 16-Aug-1996  
C/Accession: S05271; S04602  
R,Kishimoto, T.  
submitted to the EMBL Data Library, March 1989  
./Reference number: S05270

A/Cross-references: EMBL:X14584  
R/Kishimoto, T.; Okajima, H.; Okumoto, T.; Taniguchi, M.  
Nucleic Acids Res. 17, 4385, 1989  
A>Title: Nucleotide sequences of the cDNAs encoding the V-regions of H- and L-chains of  
A/Reference number: S04601; MUID:89296497; PMID:2500644

A: Residues: 1-144 <K12>  
A: Cross-references: EMBL:X14584  
C: Superfamily: Immunoglobulin V region, immunoglobulin homology  
E: Keywords: heterodimer, immunoglobulin  
F: 1/9/Domains: signal sequence #status predicted <SIS>  
F: 20-160/Product: Ig heavy chain (fragment) #status predicted <Mat>  
F: 34-117/Domains: immunoglobulin homology <IMM>

Query Match	53.1%;	Score 369;	DB 2;	Length 160;
Best Local Similarity	64.3%;	Pred. No. 1.6e-24;		
Matches	81;	Conservative	9;	Mismatches 30;
				Indels 6;
				Gaps 3;

**QY** QVQLTQSGGGLVPTGGSLRLSCAASPGIDYVMGWYQAPEKQEWMASTATGAPEN-Y-59  
::: :  
**Dd** EVQLLESGGGLVPGGSRLSCAASFRTSYLAWMWYQAPEKGLIEWASTLSGGCGSTYY-79  
20  
**QY** TDSVKGRFTISRPNDKQTEYLQNNLKPEDTAVYYCAQVRYSBSDYTN----WGCGT-11E  
60  
**Dd** ADSVKGFPTISRNSKNLTLYLNMSLRADITAVYC-AAVAVRGVTSYYIGMDWVGCGT-13B  
80

QY	116	QVTSS	121
Db	139	TVTSS	144

RESULT 8  
S31588  
Ig heavy chain V region - human (fragment)  
C:Species: Homo sapiens (man)

C:Date: 22-Nov-1993 #sequence\_revision 10-Nov-1995 #ext\_change 23-Jul-1999  
C:Accession: S31588  
R:Christier, A.M.; Gauthier, L.; Bonbli, L.; Fougereau, M.; Tonnelle, C.  
Submitted to the EMBL Data Library, June 1992  
A:Description: Mechanisms that generate human immunoglobulin diversity operate from the 8  
A:Reference number: S31585  
A:Accession: S31588  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-140 <CUI>  
A:Cross-references: EMBL:Z14200; NID:G30957; PIDN:CAA78659.1; PID:G30958  
C:Superfamily: immunoglobulin V region; immunoglobulin homology  
C:Keywords: heterotrimer; immunoglobulin  
E:34-117/Domain: immunoglobulin homology <IMM>

Query Match	52.8%;	Score 367;	DB 2;	Length 140;
Best Local Similarity	62.2%;	Pred. No. 2.1e-24;		
Matches 79;	Conservative 10;	Mismatches 26;	Indels 12;	Gaps 3;

[illegible]

Qy	115	TQVTSS	121
Db	134	TLVTSS	140

RESULT 9  
S46391

C/Date: 27-Jan-1995 #sequence\_revision 27-Jan-1995 #text\_change 20-Jun-2000  
C/Accession: S46391  
C/Entry: R/Entry: M/ Marks, J.D.; Winter, G.; Griffiths, A.D.  
J. Mol. Biol. 239, 68-78, 1994  
A/Title: In vitro assembly of repertoires of antibody chains on the surface  
A/Reference number: S46390; PMID:94254092; PMID:8196048

A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1116 <FIG>  
A:Cross-references: EMBL:Z1687; NID:G509784; PIRN:CAA8342.1, PIR:G1351444  
C:Superfamily: immunoglobulin V region, immunoglobulin homology  
C:Keywords: heterotetramer; immunoglobulin  
C:15-98/Domain: immunoglobulin homology <IM>

Query Match	52.4%	Score 364.5;	DB 2;	Length 114;
Best Local Similarity	63.9%;	Pred. No. 2.8e-24;		
Matches	78;	Conservative	8;	Mismatches 27;
			Indels	9;
			Gaps	2;

QY QVQLDQSGGGLVOTQGSFLRLSCAASGPGIDVYLWGHWQAQGRKEMWASLTATGPPN-y 59  
 : : : : :  
 Db QVTLRRSGGGLVPGGSLRLSCSASGFFSSVAHWVAQAQGLQEWAVAIWYDSKKY 60  
 1 QVTLRRSGGGLVPGGSLRLSCSASGFFSSVAHWVAQAQGLQEWAVAIWYDSKKY 60  
 QY TDSVKRFITISDNDKNTETVOMNLLKEPPAVYVCAQAQVRRSSDPTNWGQGVTV 119  
 : : : : :  
 Db ADSVKRFITISDNDKNTETVLOMNSLRQEDPAVYC-----ARDGGVGGGLVTV 112

QY	120	SS	121
Db	113	SS	114

```
RESULT 10
S20782
Ig heavy chain V region - human
CISpecies: Homo sapiens (man)
C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 23-Mar-2001
```

C/Accession: S20782  
 R/Morart, F.; Wang, J.; Schroeder, H.W.  
 submitted to the EMBL Data Library, April 1992  
 A/Description: Analysis of the IGA and IGG rearranged VH repertoire of human cord blood  
 A/Reference number: S20765  
 A/Accession: S20782  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-124 <MOR>  
 A/Cross-references: EMBL:211946; NID:G33897; PIDN:CAA78003.1; PID:G33898  
 C/Superfamily: immunoglobulin V region; immunoglobulin homology  
 C/Keywords: heterotrimer; immunoglobulin  
 F:15-98/Domain: immunoglobulin homology <IMM>

Query Match 52.4%; Score 364.5; DB 2; Length 124;  
 Best Local Similarity 62.4%; Pred. No. 36-24;  
 Matches 78; Conservative 13; Mismatches 29; Indels 5; Gaps 3;

QY 1 QVQLDSGGGLVOTGSLRLSCAASGPIGDIYIMGWYRQAPGKOREMVASITATGPPN-Y 59  
 DB 1 EVQLLESGGGLVOPGSSLRLSCAASGFTPSYAMWVRQAPGKLEWVSATISGSGSTYY 60  
 QY 60 TDSVKGRFTISRNDKNTETYLQNNLKPEDTAVYYCAQVRFRSS--DYTNWGQGTQ 116  
 DB 61 ADSVKGRFTISRNDKNTETYLQNNLSRAEDTAVYYC-AKERIAIFGVVIFHFDYWGQGT 119  
 QY 117 VTYS 121  
 DB 120 VTYS 124

RESULT 11  
 C36005  
 Ig heavy chain V region (30p1) - human  
 C/Species: Homo sapiens (man)  
 C/Date: 21-Dec-1990 #sequence\_revision 21-Dec-1990 #text\_change 16-Aug-1996  
 C/Accession: C36005  
 R/Schroeder Jr., H.W.; Wang, J.Y.  
 Proc. Natl. Acad. Sci. U.S.A. 87, 6146-6150, 1990  
 A/Title: Preferential utilization of conserved immunoglobulin heavy chain variable gene  
 A/Reference number: A36005; MUID:90349571; PMID:2117273  
 A/Accession: C36005  
 A/Status: preliminary  
 A/Molecule type: mRNA  
 A/Residues: 1-119 <SCH>  
 A/Cross-references: GB:M18513  
 C/Superfamily: immunoglobulin V region; immunoglobulin homology  
 C/Keywords: heterotrimer; immunoglobulin  
 F:15-98/Domain: immunoglobulin homology <IMM>

Query Match 52.2%; Score 363; DB 2; Length 119;  
 Best Local Similarity 63.9%; Pred. No. 3-9e-24;  
 Matches 78; Conservative 11; Mismatches 29; Indels 4; Gaps 3;

QY 1 QVQLDSGGGLVOTGSLRLSCAASGPIGDIYIMGWYRQAPGKOREMVASITATGPPN-Y 59  
 DB 1 EVQLLESGGGLVOPGSSLRLSCAASGFTPSYAMWVRQAPGKLEWVSATISGSGSTYY 60  
 QY 60 TDSVKGRFTISRNDKNTETYLQNNLKPEDTAVYYCAQVRFRSSDYTNWGQGTQ 119  
 DB 61 ADSVKGRFTISRNDKNTETYLQNNLSRAEDTAVYYCAKD--AGWGSGLF-DYWGQGT 117  
 QY 120 SS 121  
 DB 118 SS 119

RESULT 12  
 S31666  
 Ig heavy chain V region - human (fragment)  
 C/Species: Homo sapiens (man)  
 C/Date: 22-Nov-1993 #sequence\_revision 10-Nov-1995 #text\_change 23-Jul-1993  
 C/Accession: S31666

R/Cuisinier, A.M.; Gauthier, L.; Boubli, L.; Fougereau, M.; Tonnelle, C.  
 submitted to the EMBL Data Library, June 1992  
 A/Description: Mechanisms that generate human immunoglobulin diversity operate from the  
 A/Reference number: S31585  
 A/Accession: S31666  
 A/Status: preliminary  
 A/Molecule type: mRNA  
 A/Residues: 1-138 <CUI>  
 A/Cross-references: EMBL:214202; NID:G30963; PIDN:CAA78571.1; PID:G30964  
 C/Superfamily: immunoglobulin V region; immunoglobulin homology  
 C/Keywords: heterotrimer; immunoglobulin  
 F:34-117/Domain: immunoglobulin homology <IMM>

Query Match 52.2%; Score 363; DB 2; Length 138;  
 Best Local Similarity 63.1%; Pred. No. 4.6e-24;  
 Matches 77; Conservative 11; Mismatches 30; Indels 4; Gaps 2;

QY 1 QVQLDSGGGLVOTGSLRLSCAASGPIGDIYIMGWYRQAPGKOREMVASITATGPPN-Y 59  
 DB 20 EVQLLESGGGLVOPGSSLRLSCAASGFTPSYAMWVRQAPGKLEWVSATISGSGSTYY 79  
 QY 60 TDSVKGRFTISRNDKNTETYLQNNLKPEDTAVYYCAQVRFRSSDYTNWGQGTQ 119  
 DB 80 ADSVKGRFTISRNDKNTETYLQNNLSRAEDTAVYYC-AKARTGYWFDLWGQGT 136  
 QY 120 SS 121  
 DB 137 SS 138

RESULT 13  
 D36005  
 Ig heavy chain V region (M43) - human  
 C/Species: Homo sapiens (man)  
 C/Date: 21-Dec-1990 #sequence\_revision 21-Dec-1990 #text\_change 16-Dec-1998  
 C/Accession: D36005  
 R/Schroeder Jr., H.W.; Wang, J.Y.  
 Proc. Natl. Acad. Sci. U.S.A. 87, 6146-6150, 1990  
 A/Title: Preferential utilization of conserved immunoglobulin heavy chain variable gene  
 A/Reference number: A36005; MUID:90349571; PMID:2117273  
 A/Accession: D36005  
 A/Status: preliminary  
 A/Molecule type: mRNA  
 A/Residues: 1-119 <SCH>  
 A/Cross-references: GB:M34024  
 C/Genetics:  
 A/Gene: GDB:IGH@; IGHDI  
 A/Cross-references: GDB:116731; OMIM:146910  
 A/Map position: 14q32.33-14q32.33  
 C/Superfamily: immunoglobulin V region; immunoglobulin homology  
 C/Keywords: heterotrimer; immunoglobulin  
 F:15-98/Domain: immunoglobulin homology <IMM>

Query Match 52.1%; Score 362; DB 2; Length 119;  
 Best Local Similarity 62.3%; Pred. No. 4.8e-24;  
 Matches 76; Conservative 14; Mismatches 28; Indels 4; Gaps 2;

QY 1 QVQLDSGGGLVOTGSLRLSCAASGPIGDIYIMGWYRQAPGKOREMVASITATGPPN-Y 59  
 DB 1 EVQLLESGGGLVOPGSSLRLSCAASGFTPSYAMWVRQAPGKLEWVSATISGSGSTYY 60  
 QY 60 TDSVKGRFTISRNDKNTETYLQNNLKPEDTAVYYCAQVRFRSSDYTNWGQGTQ 119  
 DB 61 ADSVKGRFTISRNDKNTETYLQNNLSRAEDTAVYYC-AKMDNDNWDPMQGT 117  
 QY 120 SS 121  
 DB 118 SS 119

RESULT 14  
 F36005  
 Ig heavy chain V region (M49) - human

Tue Oct 5 09:51:21 2004

us-09-805-290a-24.rpr

Page 5

Search completed: October 5, 2004, 08:17:16  
Job time : 22.0648 secs

C/Species: Homo sapiens (man)  
C/Date: 21-Dec-1990 #sequence\_revision 21-Dec-1990 #text\_change 16-Dec-1998  
C/Accession: F36005  
R/Schroeder Jr., H.W.; Wang, J.Y.  
Proc. Natl. Acad. Sci. U.S.A. 87, 6146-6150, 1990  
A/Title: Preferential utilization of conserved immunoglobulin heavy chain variable gene  
A/Reference number: A36005; MUID:90349571; PMID:2117273  
A/Accession: F36005  
A/Status: preliminary  
A/Molecule type: mRNA  
A/Residues: 1-119 <SCH>  
A/Cross-references: GB:M34026  
C/Genetics:  
A/Gene: GDB:IGH@; IGHDI1  
A/Cross-references: GDB:118731; OMIM:146910  
A/Map position: 14q32.33-14q32.33  
C/Superfamily: immunoglobulin V region; immunoglobulin homology  
C/Keywords: heterotrimer; immunoglobulin  
F/15-98/Domain: immunoglobulin homology <IMM>

Query Match 52.1%; Score 362; DB 2; Length 119;  
Best Local Similarity 65.6%; Pred. No. 4.8e-24;  
Matches 80; Conservative 7; Mismatches 31; Indels 4; Gaps 2;

QY 1 QVQLDSSGGGLVQPGGSLRLSCAASGPIGDIYLMGWYRQAPGKOREMVASITATGPPN-Y 59  
DB 1 QVQLVESGGGVQPGSRSLRLSCAASGFTSSYAMHWRAQAPGKLEWYAVISYDGSNKYY 60  
QY 60 TDSVKGRTTISRNDKNTREYLOMNNLKPEDPAVYYCNAQVRRPSDYTNWGGTQVTV 119  
DB 61 ADSVKGRTTISRDNKNTLYLOMNSLRADPAVYYC---ARDRKASDAFDIWGQITVTV 117  
QY 120 SS 121  
DB 118 SS 119

# RESULT 15

1g heavy chain - human  
S31112  
C/Species: Homo sapiens (man)  
C/Date: 02-Dec-1993 #sequence\_revision 26-May-1995 #text\_change 17-Mar-1999  
C/Accession: S31112  
R/Raaphorst, F.M.; Timmers, E.; Kentler, M.J.H.; van Tol, M.J.D.; Vossen, J.M.; Schuurman  
Eur. J. Immunol. 22, 247-251, 1992  
A/Title: Restricted utilization of germ-line V(H)3 genes and short diverse third complement  
A/Reference number: S31104; MUID:92111633; PMID:1730252  
A/Accession: S31112  
A/Status: preliminary; nucleic acid sequence not shown; translation not shown  
A/Molecule type: mRNA  
A/Residues: 1-120 <RAA>  
A/Cross-references: EMBL:X62961  
A/Note: The nucleotide sequence was submitted to the EMBL Data Library, October 1991  
C/Superfamily: immunoglobulin V region; immunoglobulin homology  
C/Keywords: heterotrimer; immunoglobulin  
F/15-98/Domain: immunoglobulin homology <IMM>

Query Match 52.0%; Score 361.5; DB 2; Length 120;  
Best Local Similarity 65.6%; Pred. No. 5.3e-24;  
Matches 80; Conservative 8; Mismatches 31; Indels 3; Gaps 2;

QY 1 QVQLDSSGGGLVQPGGSLRLSCAASGPIGDIYLMGWYRQAPGKOREMVASITATGPPN-Y 59  
DB 1 QVQLVESGGGVQPGSRSLRLSCAASGFTSSYAMHWRAQAPGKLEWYAVISYDGSNKYY 60  
QY 60 TDSVKGRTTISRNDKNTREYLOMNNLKPEDPAVYYCNAQVRRPSDYTNWGGTQVTV 119  
DB 61 ADSVKGRTTISRDNKNTLYLOMNSLRADPAVYYCATGVVVVAA--TDYWGQITVTV 118  
QY 120 SS 121  
DB 119 SS 120

This page Blank (uspio)



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2004, 07:29:35 : Search time 11.6358 Seconds  
(without alignments)  
581.749 Million cell updates/sec

Title: US-09-805-290A-24

Perfect score: 695  
Sequence: 1 QVQLQDSGGGLVQTGSLRL.....WGQGTQVTVSSSEPKTKPPQ 130

Scoring table: BIOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	340.5	49.0	126 1 HV3X_HUMAN	P01772 homo sapien
2	340	48.9	115 1 HV3D_HUMAN	P01766 homo sapien
3	338	48.6	119 1 HV3I_HUMAN	P01770 homo sapien
4	335.5	48.3	122 1 HV3G_HUMAN	P01768 homo sapien
5	323	46.5	115 1 HV3F_HUMAN	P01767 homo sapien
6	322.5	46.4	116 1 HV3T_HUMAN	P01810 mus musculu
7	322.5	46.4	119 1 HV4O_MOUSE	P01810 mus musculu
8	322	46.3	121 1 HV3J_HUMAN	P01771 homo sapien
9	321.5	46.3	119 1 HV3C_HUMAN	P01764 mus musculu
10	320.5	46.1	117 1 HV3C_HUMAN	P01764 mus musculu
11	319.5	46.0	114 1 HV3B_HUMAN	P01763 homo sapien
12	318.5	45.8	122 1 HV3A_HUMAN	P01762 homo sapien
13	317	45.6	115 1 HV33_MOUSE	P01802 mus musculu
14	317	45.6	120 1 HV3H_HUMAN	P01766 homo sapien
15	316.5	45.5	122 1 HV3E_HUMAN	P01769 mus musculu
16	313.5	45.1	114 1 HV01_CANFA	P01784 canis famli
17	313.5	45.1	136 1 HV16_MOUSE	P01783 mus musculu
18	313	45.0	113 1 HV30_MOUSE	P01799 mus musculu
19	313	45.0	115 1 HV32_MOUSE	P01801 mus musculu
20	312	44.9	118 1 HV39_MOUSE	P01809 mus musculu
21	310.5	44.7	119 1 HV38_MOUSE	P01808 mus musculu
22	309	44.5	116 1 HV05_CARAU	P19181 carassius a
23	309	44.5	117 1 HV02_CANFA	P01785 canis famli
24	308.5	44.4	120 1 HV3U_HUMAN	P01782 homo sapien
25	308	44.3	113 1 HV27_MOUSE	P01796 mus musculu
26	308	44.3	119 1 HV3L_HUMAN	P01773 homo sapien
27	303.5	43.7	119 1 HV3N_HUMAN	P01775 homo sapien
28	302.5	43.5	111 1 HV35_MOUSE	P01804 mus musculu
29	302	43.5	113 1 HV28_MOUSE	P01797 mus musculu
30	302	43.5	142 1 HV01_RAT	P01805 rattus norv
31	301	43.3	117 1 HV42_MOUSE	P01812 mus musculu
32	299.5	43.1	119 1 HV3P_HUMAN	P01777 homo sapien
33	299	43.0	113 1 HV3I_MOUSE	P01800 mus musculu

## ALIGNMENTS

RESULT 1	ID	HV3X_HUMAN	STANDARD;	PRT;	126 AA.
AC P01772;					
DT 21-JUL-1986 (Rel. 01, Created)					
DT 21-JUL-1986 (Rel. 01, Last sequence update)					
DT 10-OCT-2003 (Rel. 42, Last annotation update)					
DE Ig heavy chain V-II region KOL.					
OS Homo sapiens (Human).					
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.					
OX NCBI_Taxid=9606;					
RN [1]					
RP SEQUENCE AND DISULFIDE BONDS.					
RX MEDLINE=63289131; PubMed=6884994;					
RA Schmitt W.E., Jung H.-D., Palm W., Hilschmann N.;					
RT "Three-dimensional structure determination of antibodies. Primary structure of crystallized monoclonal immunoglobulin IgG1 KOL, I.";					
RL Hoppe-Seyler's Z. Physiol. Chem. 364:713-747(1983).					
CC 1- SIMILARITY: Contains 1 immunoglobulin-like domain.					
CC PIR; A02055; GIHQUL.					
DR PDB; 2FB4; 12-JUL-89.					
DR GO; GO:000576; C:extracellular; NAS.					
DR GO; GO:0003823; F:antigen binding; NAS.					
DR GO; GO:0006955; P:immune response; NAS.					
DR InterPro; IPR007110; Ig-like.					
DR InterPro; IPR003596; Ig_V.					
DR Pfam; PF00047; Ig_1.					
DR SMART; SM00406; IgV_1.					
DR PROSITE; PSS0835; IG_LIKE; 1.					
KM Immunoglobulin V region; 3D-structure; Pyrrolidone carboxylic acid.					
FT DOMAIN 1 112					
FT MOD_RES 1 1					
FT DISULFID 22 96					
FT DISULFID 105 110					
FT STRAND 3 7					
FT STRAND 11 12					
FT TURN 14 15					
FT STRAND 18 25					
FT STRAND 29 31					
FT HELIX 34 39					
FT STRAND 41 42					
FT TURN 45 51					
FT STRAND 53 54					
FT TURN 58 60					
FT STRAND 62 64					
FT HELIX 65 65					
FT STRAND					

FT TURN 66 67  
 FT STRAND 68 73  
 FT TURN 74 77  
 FT STRAND 78 83  
 FT HELIX 88 90  
 FT STRAND 92 99  
 FT STRAND 106 108  
 FT TURN 107 108  
 FT STRAND 109 109  
 FT STRAND 113 116  
 FT STRAND 120 124  
 FT NON TER 126 126  
 SQ SEQUENCE 126 AA; 13718 MW; E4D71B52B16F0776 CRC64;

Query Match 49.0%; Score 340.5; DB 1; Length 126;  
 Best Local Similarity 59.5%; Pred. No. 4.6e-26;  
 Matches 75; Conservative 12; Mismatches 34; Indels 5; Gaps 2;

QY 1 QVQLDSSGGGVVOTGGSLRLSCAASGPGIDVYLMGWYRQAPKQREEMVASITATG-PNNY 59  
 DB 1 QVQLVSGGGVQVQPGKSLRLSCSSGFTFSSTAYMTWVQAPKGLWVAITWDDSDQHY 60  
 QY 60 TDSVKGRTISRDNKNTLEYLQNNMLKPEDTAVYYCAQVRVRFSSDYT---NYWGQGT 115  
 DB 61 ADSVKGRFTISRDNKNTLFLQMDSLRPEDTGVYFCARDGKHGFCSSASCRGPDYWGQGT 120  
 QY 116 QVTVSS 121  
 DB 121 PVTVSS 126

## RESULT 2

HV3D\_HUMAN STANDARD; PRT; 115 AA.  
 ID HV3D\_HUMAN

AC P01765;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Ig heavy chain V-II region TIL.  
 OS Homo sapiens (Human)  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 ON NCBI\_TaxID=9606;  
 RX MEDLINE=78005528; PubMed=409716;  
 RP SEQUENCE.  
 RA Wang A.-C., Wang I.Y., Fudenberg H.H.;  
 RT "Immunoglobulin structure and genetics. Identity between variable  
 RT regions of a mu and a gamma2 chain."  
 RL J. Biol. Chem. 252:7192-7199 (1977).  
 CC -1- MISCELLANEOUS: THE SEQUENCES OF THE V REGIONS OF THE HEAVY CHAINS  
 CC OF IGM AND IGG2 ISOLATED FROM A SINGLE PATIENT WITH BICLONAL  
 CC GAMMOPATHY ARE IDENTICAL. THEIR LIGHT CHAINS ARE APPARENTLY ALSO  
 CC IDENTICAL.  
 CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.  
 DR PIR; A02048; H3HUTL.  
 DR HSSP; P01772; 2F84.  
 DR GO; GO:0005576; C:extracellular; NAS.  
 DR GO; GO:0003823; P:antigen binding; NAS.  
 DR GO; GO:0006955; P:immune response; NAS.  
 DR InterPro; IPR007110; Ig-like.  
 DR InterPro; IPR003596; Ig\_V.  
 DR Pfam; PF00047; Ig\_1.  
 DR SMART; SMO0406; IGV; 1.  
 DR PROSITE; PS50835; IG LIKE; 1.  
 DR Immunoglobulin V region.  
 DR DOMAIN 1 108 IG-LIKE.  
 FT NON TER 115 115  
 SQ SEQUENCE 115 AA; 12356 MW; 4DC67D179FE2326 CRC64;

Query Match 48.9%; Score 340; DB 1; Length 115;  
 Best Local Similarity 58.2%; Pred. No. 4.6e-26;  
 Matches 71; Conservative 15; Mismatches 28; Indels 8; Gaps 3;

QY 1 QVQLDSSGGGVVOTGGSLRLSCAASGPGIDVYLMGWYRQAPKQREEMVASITATG-PNNY 59  
 DB 1 EVQLLESGGGLVQPGKSLRLSCAASGFTFSSTYMTWVQAPKGLZMWVAITZLSYSSZY 60  
 QY 60 TDSVKGRTISRDNKNTLEYLQNNMLKPEDTAVYYCAQVRVRFSSDYNYWGQGTQVTV 119  
 DB 61 ABSVKGRFTISRDNKNTLFLQMDSLRPEDTGVYFCARDGKHGFCSSASCRGPDYWGQGT 113  
 QY 120 SS 121  
 DB 114 SS 115

## RESULT 3

HV3I\_HUMAN STANDARD; PRT; 119 AA.  
 ID HV3I\_HUMAN

AC P01770;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Ig heavy chain V-III region NIE.  
 OS Homo sapiens (Human)  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 ON NCBI\_TaxID=9606;  
 RX MEDLINE=77070269; PubMed=826475;  
 RP SEQUENCE.  
 RA Postelng H., Hilschmann N.;  
 RT "The rule of antibody structure. The primary structure of a  
 RT monoclonal IgG1 immunoglobulin (myeloma protein NIE). III. The  
 RT cytotryptic peptides of the H-chain, alignment of the tryptic  
 RT peptides and discussion of the complete structure."  
 RL Hoppe-Seyler's Z. Physiol. Chem. 357:1571-1604 (1976).  
 RN [2]  
 RP DISULFIDE BOND.  
 RX MEDLINE=77070267; PubMed=1002129;  
 RA Dreker L., Schwarz J., Reichel W., Hilschmann N.;  
 RT "Rule of antibody structure. The primary structure of a monoclonal  
 RT IgG1 immunoglobulin (myeloma protein NIE). I: Purification and  
 RT characterization of the protein, the L- and H-chains, the  
 RT cyanogen bromide cleavage products, and the disulfide bridges."  
 RL Hoppe-Seyler's Z. Physiol. Chem. 357:1515-1540 (1976).  
 CC -1- MISCELLANEOUS: THIS CHAIN WAS ISOLATED FROM AN IGG1 MYELOMA  
 CC PROTEIN.  
 CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.  
 DR PIR; A91668; GIHUNI.  
 DR HSSP; P01772; 2F84.  
 DR GO; GO:0005576; C:extracellular; NAS.  
 DR GO; GO:0003823; P:antigen binding; NAS.  
 DR GO; GO:0006955; P:immune response; NAS.  
 DR InterPro; IPR007110; Ig-like.  
 DR InterPro; IPR003596; Ig\_V.  
 DR Pfam; PF00047; Ig\_1.  
 DR SMART; SMO0406; IGV; 1.  
 DR PROSITE; PS50835; IG LIKE; 1.  
 DR Immunoglobulin V region; Pyrolydine carboxylic acid.  
 DR DOMAIN 1 112 IG-LIKE.  
 FT MOD RES 1 1  
 FT DISULFID 22 96 PYRROLIDONE CARBOXYLIC ACID.  
 FT NON TER 119 119  
 SQ SEQUENCE 119 AA; 13242 MW; C96935A6B5E165B CRC64;

Query Match 48.6%; Score 338; DB 1; Length 119;  
 Best Local Similarity 59.8%; Pred. No. 7.5e-26;  
 Matches 73; Conservative 14; Mismatches 31; Indels 4; Gaps 2;  
 QY 1 QVQLDSSGGGVVOTGGSLRLSCAASGPGIDVYLMGWYRQAPKQREEMVASITATG-PNNY 59  
 DB 1 QVQLVSGGGVQVQPGKSLRLSCAASGFTFSRTYIMVQAPKGLWVAIVMSYBGBKHV 60  
 QY 60 TDSVKGRTISRDNKNTLEYLQNNMLKPEDTAVYYCAQVRVRFSSDYNTNWGQGTQVTV 119

```

Db 61 ADVNAGFTTISRDNSKNTLYLNMSLRPEDTAVYVC---ARIDXTMFPAMQGLTIV 117
QY 120 SS 121
Db 118 SS 119

RESULT 4
HV3F HUMAN STANDARD; PRT; 122 AA.
ID HV3F HUMAN
AC P01768;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ig heavy chain V-II region CAM.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE.
RA MEDLINE=81013859; PubMed=6774332;
RA Lehman D.W., Putnam F.W.;
RT "Amino acid sequence of the variable region of a human mu chain:
RT location of a possible JH segment."
RL Proc. Natl. Acad. Sci. U.S.A. 77:3239-3243(1980).
CC -1- MISCELLANEOUS: THIS MU CHAIN WAS ISOLATED FROM THE PLASMA OF A
CC PATIENT WITH MACROGLOBULINEMIA.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.
DR PIR: A02051; M3HUM.
DR HSSP: P01772; 2FB4.
DR GO: GO:0005576; C:extracellular; NAS.
DR GO: GO:0003823; P:antigen binding; NAS.
DR GO: GO:0006955; P:immune response; NAS.
DR InterPro: IPR007110; IG-like.
DR InterPro: IPR003596; IG_V.
DR Pfam: PF00047; Ig_1.
DR SMART: SM00406; IGV; 1.
DR PROSITE: PS00835; IG LIKE; 1.
KM Immunoglobulin V region; Pyrrrolidone carboxylic acid.
FT DOMAIN 1 112
FT MOD_RES 1 112
FT NON_TER 1 112
SQ SEQUENCE 122 AA; 13668 MW; A42D0F17D252F1C2 CRC64;

Query Match 48.3%; Score 335.5; DB 1; Length 122;
Best Local Similarity 59.7%; Pred. No. 1.3e-25;
Matches 74; Conservative 14; Mismatches 31; Indels 5; Gaps 4;

QY 1 QVQLQDSGGGLVQTGSLRLSCAASGPIGDIVLNGWYRQAPGKQREMYVASITATGPPN-Y 59
Db 1 QVELVESGGGVYVZPSRSLRLSCAASGFTFSNVAMHWVQPGKGLIEWAVISYGBBXY 60
QY 60 TDSVYGRFTISRNDKNTLEYLQNMNLKPEDTAVYVCNAQVRFSSDXT--NWGGQGV 117
Db 61 ABSYVGRFTISRDSKNTLYLNMSLRABDTAVYVC-ARDPPLV-GBYRAFNWGGGLTV 118
QY 118 TVSS 121
Db 119 TVSS 122

RESULT 5
HV3F HUMAN STANDARD; PRT; 115 AA.
ID HV3F HUMAN
AC P01767;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ig heavy chain V-II region BUI.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

```

OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE.
RA MEDLINE=78137069; PubMed=416441;
RA Torano A., Putnam F.W.;
RT "Complete amino acid sequence of the alpha 2 heavy chain of a human
RT IgA2 immunoglobulin of the A2m (2) allotype."
RL Proc. Natl. Acad. Sci. U.S.A. 75:966-969(1978).
CC -1- MISCELLANEOUS: THE SEQUENCE OF THE ALPHA-2, A2M(2) ALLOTYPE, C
CC REGION OF THIS MYELOMA PROTEIN IS ALSO GIVEN.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.
DR PIR: A02050; A2HUBU.
DR HSSP: P01789; IMCP.
DR GO: GO:0005576; C:extracellular; NAS.
DR GO: GO:0003823; P:antigen binding; NAS.
DR GO: GO:0006955; P:immune response; NAS.
DR InterPro: IPR007110; IG-like.
DR InterPro: IPR003596; IG_V.
DR Pfam: PF00047; Ig_1.
DR SMART: SM00406; IGV; 1.
DR PROSITE: PS00835; IG LIKE; 1.
KM Immunoglobulin V region.
FT DOMAIN 1 111
FT NON_TER 1 115
SQ SEQUENCE 115 AA; 12379 MW; 208876A7DF52DCP4 CRC64;

Query Match 46.5%; Score 323; DB 1; Length 115;
Best Local Similarity 54.5%; Pred. No. 2e-24;
Matches 66; Conservative 17; Mismatches 32; Indels 6; Gaps 1;

QY 1 QVQLQDSGGGLVQTGSLRLSCAASGPIGDIVLNGWYRQAPGKQREMYVASITATGPPN-Y 60
Db 1 EVQLVETGGGLTLPGGSLRLSCAASGFTVSHSMGHWVRQAPGKALZWVSATIRGGTTYA 60
QY 61 DSVYGRFTISRNDKNTLEYLQNMNLKPEDTAVYVCNAQVRFSSDXTNWGGQGVTVS 120
Db 61 DSVYGRFTISRDSKNTLYLNMSLRABDTAVYVCARDLAA-----ARLPKQITTVYS 114
QY 121 S 121
Db 115 S 115

RESULT 6
HV3T HUMAN STANDARD; PRT; 116 AA.
ID HV3T HUMAN
AC P01781;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ig heavy chain V-II region GAL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE.
RA MEDLINE=75059123; PubMed=4803843;
RA Watanabe S., Barinkol H.U., Horn J., Bertram J., Hilschmann N.;
RT "The primary structure of a monoclonal IgM-immunoglobulin
RT (macroglobulin Gal.), II: the amino acid sequence of the H-chain (mu-
RT type), subgroup H III. Architecture of the complete IgM-molecule."
RL Hoppe-Seyler's Z. Physiol. Chem. 354:1505-1509(1973).
RN [2]
RP REVISION TO 28-93.
RA Hilschmann N.;
RL Submitted (JUN-1975) to the PIR data bank.
CC -1- MISCELLANEOUS: THIS MU CHAIN WAS ISOLATED FROM A WALDENSTROM'S
CC MACROGLOBULIN.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.
DR PIR: A02064; M3HUGL.
DR HSSP: P01772; 2FB4.

```

DR GO; GO:0005576; C:extracellular; NAS.  
 DR GO; GO:0003823; F:antigen binding; NAS.  
 DR GO; GO:0006955; P:immune response; NAS.  
 DR InterPro; IPR007110; IG-like.  
 DR InterPro; IPR003596; IG\_v.  
 DR Pfam; PF00047; IG\_1.  
 DR SMART; SM00406; IGV; 1.  
 DR PROSITE; PS00835; IG-LIKE; 1.  
 KW Immunoglobulin V region.  
 FT DOMAIN 1 112 IG-LIKE.  
 FT NON TER 116 116  
 SQ SEQUENCE 116 AA; 12730 MW; 2C67CA9AAAAA1282 CRC64;

Query Match 46.4%; Score 322.5; DB 1; Length 116;  
 Best Local Similarity 60.7%; Pred. No. 2.3e-24;  
 Matches 74; Conservative 10; Mismatches 31; Indels 7; Gaps 3;

QY 1 QVQLDPSGGLVOTGSLRLSCAASGIPDGYLIMGVYRQAPGKQEMVASIATG-PNY 59  
 DB 1 EVQLVSGGGLVQPGSLRLSCAASGIPDGYLIMGVYRQAPGKQEMVASIATGSSZZBY 60

QY 60 TDSVKGKFTISRNDKNTLEYLQNNIKPEDTAVYYCNAQVRVRFSSDYTNWGQGTQVTV 119  
 DB 61 VDSVKGKFTISRDNANSLYLQNNISRVEDTALYYC---ARFGWGGLD---YWGQGTQVTV 114

QY 120 SS 121  
 DB 115 ST 116

Db 115 ST 116

RESULT 7  
 HV40\_MOUSE STANDARD; PRT; 119 AA.  
 ID HV40\_MOUSE  
 AC P01810;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE IG heavy chain V region J539.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OC NCBI\_TaxID=10090;  
 RN [1]  
 RP PRELIMINARY SEQUENCE.  
 RX MEDLINE=79233895; PubMed=11245;  
 RA Rao D.N., Rudikoff S., Kruttsch H., Potter M.;  
 RT "Structural evidence for independent joining region gene in  
 immunoglobulin heavy chains from anti-galactan myeloma proteins and  
 its potential role in generating diversity in  
 complementarity-determining regions.";  
 RT Proc. Natl. Acad. Sci. U.S.A. 76:2890-2894(1979).  
 RL [2]  
 RP X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS).  
 RX MEDLINE=88217852; PubMed=3449653;  
 RA Suh S.W., Bhat T.N., Navia M.A., Cohen G.H., Rao D.N., Rudikoff S.,  
 RA Davies D.R.;  
 RT "The galactan-binding immunoglobulin Fab J539: an X-ray diffraction  
 study at 2.6-A resolution.";  
 RT Proteins 1:74-80(1986).  
 CC -1- MISCELLANEOUS: THIS CHAIN WAS ISOLATED FROM A MYELOMA PROTEIN THAT  
 BINDS GALACTAN.  
 CC PIR: A02080; AVMSJ5.  
 DR PDB: 2FB7; 15-OCT-90.  
 DR InterPro; IPR007110; IG-like.  
 DR InterPro; IPR003596; IG\_v.  
 DR Pfam; PF00047; IG\_1.  
 DR SMART; SM00406; IGV; 1.  
 DR PROSITE; PS00835; IG-LIKE; 1.  
 KW Immunoglobulin V region; 3D-structure.  
 FT NON TER 119 119  
 FT STRAND 3 7  
 FT STRAND 10 12  
 FT TURN 14 15

FT STRAND 18 25  
 FT HELIX 29 31  
 FT STRAND 34 39  
 FT TURN 41 42  
 FT STRAND 45 51  
 FT TURN 53 54  
 FT STRAND 58 60  
 FT TURN 62 67  
 FT STRAND 68 72  
 FT STRAND 78 83  
 FT HELIX 88 90  
 FT STRAND 92 100  
 FT TURN 101 103  
 FT STRAND 104 108  
 FT STRAND 112 116  
 SQ SEQUENCE 119 AA; 13240 MW; 57B4F1DB675C1F1 CRC64;

Query Match 46.4%; Score 322.5; DB 1; Length 119;  
 Best Local Similarity 56.1%; Pred. No. 2.4e-24;  
 Matches 69; Conservative 16; Mismatches 33; Indels 5; Gaps 3;

QY 1 QVQLDPSGGLVOTGSLRLSCAASGIPDGYLIMGVYRQAPGKQEMVASI-TATGPPNY 59  
 DB 1 EVQLVSGGGLVQPGSLRLSCAASGIPDGYLIMGVYRQAPGKQEMVASIHPDGGTIN 60

QY 60 TDSVKGKFTISRNDKNTLEYLQNNIKPEDTAVYYCNAQVRVRFSSDYTNWGQGTQVTV 119  
 DB 61 TDSVKGKFTISRDNANSLYLQNNISRVEDTALYYC---ARLHYG-YNAVWGQGTQVTV 116

QY 120 SSE 122  
 DB 117 SAE 119

Db 117 SAE 119

RESULT 8  
 HV30\_HUMAN STANDARD; PRT; 121 AA.  
 ID HV30\_HUMAN  
 AC P01771;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE IG heavy chain V-III region H1L.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OC NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE.  
 RX MEDLINE=79124695; PubMed=420800;  
 RA Chiu Y.-Y.H., Lopez de Castro J.A., Poljak R.J.;  
 RT "Amino acid sequence of the VH region of human myeloma  
 RT cryoimmunoglobulin IgG H1L.";  
 RT Biochemistry 18:553-560(1979).  
 RL [2]  
 CC -1- MISCELLANEOUS: THIS CHAIN WAS ISOLATED FROM AN IGGI MYELOMA  
 CC PROTEIN.  
 CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.  
 DR PIR: A02054; GIHJHL.  
 DR HSSP; P01772; 2FB4.  
 DR GO; GO:0005576; C:extracellular; NAS.  
 DR GO; GO:0003823; F:antigen binding; NAS.  
 DR GO; GO:0006955; P:immune response; NAS.  
 DR InterPro; IPR007110; IG-like.  
 DR InterPro; IPR003596; IG\_v.  
 DR Pfam; PF00047; IG\_1.  
 DR SMART; SM00406; IGV; 1.  
 DR PROSITE; PS00835; IG-LIKE; 1.  
 KW Immunoglobulin V region; Pyroliidone carboxylic acid.  
 FT DOMAIN 1 112 IG-LIKE.  
 FT MOD\_RES 1 1 PYROLIDONE CARBOXYLIC ACID.  
 FT NON TER 121 121  
 SQ SEQUENCE 121 AA; 13566 MW; 480FC53610EF5DAB CRC64;

Query Match 46.3%; Score 322; DB 1; Length 121;

Best Local Similarity: 58.9%; Pred. No. 2.7e-24;

Matches 73; Conservative 9; Mismatches 36; Indels 6; Gaps 3;

```

QY 1 QVQLDSSGGGLVQPGGSLRLSCAASGPIGDIYLMGWYRQAPGKOREMVASITATGPPN-Y 59
DB 1 QVQLVQAGGGVQPGGSLRLSCISAGFIFSNYGMHWYQAQGGKLEWVAIVMGSRITY 60
QY 60 TDSVKGRTISRDNDKNTETYLQNNLKPEDTAVYYC--NAGVRRFSSDYTNWGQGTQV 117
DB 61 GDSVKGRTISRDNDKNTETYLQNNLKPEDTAVYYCARDPDLTAFSD--YWGQGLTV 117
QY 118 TVSS 121
DB 118 TVSS 121
DB 118 TVSS 121

```

## RESULT 9

```

HV37_MOUSE STANDARD; PRT; 119 AA.
ID HV37_MOUSE
AC P01807;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ig heavy chain V region X44.
OS Mus musculus (Mouse).
CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CX NCBI_Taxid=10090;
RN [1]

```

## SEQUENCE.

```

RX MEDLINE=79223895; PubMed=111245;
RA Rao D.N., Rudikoff S., Krutzsch H., Potter M.;
RT "Structural evidence for independent joining region gene in
RT immunoglobulin heavy chains from anti-galactan myeloma proteins and
RT its potential role in generating diversity in
RT complementarity-determining regions."
RL Proc. Natl. Acad. Sci. U.S.A. 76:2890-2894(1979).
CC -1- MISCELLANEOUS: THIS CHAIN WAS ISOLATED FROM AN IGA MYELOMA PROTEIN
CC THAT BINDS GALACTAN.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.
DR PIR; A02077; AWMGX4.
DR HSSP; P01810; 2PRF.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003596; Ig_V.
DR Pfam; PF00047; Ig_1.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS50835; IG-LIKE; 1.
KW Immunoglobulin V region.
FT DOMAIN 1
FT NON_TER 119
FT SEQUENCE 119 AA; 13246 MW; BC34FC8P31CD41B3 CRC64;

```

Query Match 46.3%; Score 321.5; DB 1; Length 119;

Best Local Similarity 56.1%; Pred. No. 2.9e-24;

Matches 69; Conservative 15; Mismatches 34; Indels 5; Gaps 3;

```

QY 1 QVQLDSSGGGLVQPGGSLRLSCAASGPIGDIYLMGWYRQAPGKOREMVASITATGPPN-Y 59
DB 1 EVKLESGGGLVQPGGSLRLSCAASGPIGDIYLMGWYRQAPGKLEWVAIVMGSRITY 60
QY 60 TDSVKGRTISRDNDKNTETYLQNNLKPEDTAVYYC--NAGVRRFSSDYTNWGQGTQV 119
DB 61 TDSVKGRTISRDNDKNTETYLQNNLKPEDTAVYYC--NAGVRRFSSDYTNWGQGTQV 119
QY 120 SSE 122
DB 120 SSE 122
QY 117 SAE 119
DB 117 SAE 119

```

```

RESULT 10
HV3C_HUMAN STANDARD; PRT; 117 AA.
ID HV3C_HUMAN
AC P01764;

```

21-JUL-1986 (Rel. 01, Created)

21-JUL-1986 (Rel. 01, Last sequence update)

10-OCT-2003 (Rel. 42, Last annotation update)

Ig heavy chain V-III region VH26 precursor.

Homo sapiens (Human).

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

NCBI\_Taxid=9606;

SEQUENCE FROM N.A.

MEDLINE=8101090; PubMed=6450418;

Mathysens G., Rabbits T.H.;

"Structure and multiplicity of genes for the human immunoglobulin

heavy chain variable region."

Proc. Natl. Acad. Sci. U.S.A. 77:6561-6565(1980).

-1- SIMILARITY: Contains 1 immunoglobulin-like domain.

-----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb.ch/announce/>

CC or send an email to [license@isb.slb.ch](mailto:license@isb.slb.ch)).

-----

DR EMBL; J00236; AA53516.1; -

DR EMBL; M35415; AA58735.1; -

DR PIR; A02047; H3H026.

DR PDB; 1H0U; 23-DEC-99.

DR Genew; H0NC5545; IGHV@.

DR GO; GO:000576; C:extracellular; NAS.

DR GO; GO:0003823; F:antigen binding; NAS.

DR GO; GO:0006955; P:immune response; NAS.

DR InterPro; IPR007110; Ig-like.

DR InterPro; IPR003596; Ig\_V.

DR Pfam; PF00047; Ig\_1.

DR SMART; SM00406; IGV; 1.

DR PROSITE; PS50835; IG-LIKE; 1.

KW Immunoglobulin V region; Signal, 3D-structure.

FT SIGNAL 1

FT CHAIN 20 117 IG HEAVY CHAIN V-III REGION VH26.

FT DOMAIN 20 >117 IG-LIKE.

FT NON\_TER 117 117

FT SEQUENCE 117 AA; 12582 MW; E826732P1A3CB0F1 CRC64;

Query Match 46.1%; Score 320.5; DB 1; Length 117;

Best Local Similarity 68.8%; Pred. No. 3.6e-24;

Matches 66; Conservative 8; Mismatches 21; Indels 1; Gaps 1;

```

QY 1 QVQLDSSGGGLVQPGGSLRLSCAASGPIGDIYLMGWYRQAPGKOREMVASITATGPPN-Y 59
DB 20 EVQLLESGLVQPGGSLRLSCAASGPIGDIYLMGWYRQAPGKLEWVAIVMGSRITY 79
QY 60 TDSVKGRTISRDNDKNTETYLQNNLKPEDTAVYYC 95
DB 80 GDSVKGRTISRDNDKNTETYLQNNLKPEDTAVYYC 115

```

## RESULT 11

```

HV3B_HUMAN STANDARD; PRT; 114 AA.
ID HV3B_HUMAN
AC P01763;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ig heavy chain V-III region VEA.
OS Homo sapiens (Human).
CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
CX NCBI_Taxid=9606;
RN [1]

```

```

RP SEQUENCE
RX MEDLINE=83273707; PubMed=6410398;

```

RA Goni F., Frangione B.;  
 RT "Amino acid sequence of the Fv region of a human monoclonal IgM  
 (protein W3) with antibody activity against 3,4-pyruvylated  
 RT galactose in Klebsiella polysaccharides K30 and K33."  
 RL Proc. Natl. Acad. Sci. U.S.A. 80:4837-4841(1983).  
 CC -1- MISCELLANEOUS: THIS CHAIN WAS OBTAINED FROM A MONOCLONAL ANTIBODY  
 AGAINST 3,4-PYRUVYLATED GALACTOSE AND ISOLATED FROM A PATIENT WITH  
 WALDENSTROM'S MACROGLOBULINEMIA.  
 CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.  
 DR PIR; A02046; M3HWE.  
 DR HSSP; P01772; 2F84.  
 DR GO; GO:0005576; Cxcracellular; NAS.  
 DR GO; GO:0003823; Piantigen binding; NAS.  
 DR GO; GO:0006955; P:immune response; NAS.  
 DR InterPro; IPR007110; Ig-like.  
 DR InterPro; IPR003596; Ig\_v.  
 DR Pfam; PF00047; Ig; 1.  
 DR SMART; SM00406; IGV; 1.  
 DR PROSITE; PSS0835; IG LIKE; 1.  
 KM Immunoglobulin V region; Pyrolydione carboxylic acid.  
 FT DOMAIN 1 112 IG-LIKE.  
 FT MOD\_RES 1 1 PYRROLIDONE CARBOXYLIC ACID.  
 FT NON\_TER 114 114  
 SQ SEQUENCE 114 AA; 12256 MW; D88294FB418A07B7 CRC64;  
 Query Match 46.0%; Score 319.5; DB 1; Length 114;  
 Best Local Similarity 57.4%; Pred. No. 4,4e-24;  
 Matches 70; Conservative 13; Mismatches 30; Indels 9; Gaps 2;  
 QY 1 QVQLDPSGGGLVQVGTGSLRLSCAASGPGIDVYIMGMYRQAPKQREMVASITATGPP-NY 59  
 DB 1 QVQLVDSGGGLVQVGTGSLRLSCAASGPGIDVYIMGMYRQAPKQREMVASITATGPP-NY 60  
 QY 60 TDSVKGKFTISRNDKNTETYLQWNLKPEPTAVYYCAQVRFSSDYTNWGQGTQVTV 119  
 DB 61 ADSVKGKFTISRNDKNTETYLQWNLKPEPTAVYYCAQVRFSSDYTNWGQGTQVTV 112  
 QY 120 SS 121  
 DB 113 SS 114  
 RESULT 12  
 HVA3\_HUMAN STANDARD; PRT; 122 AA.  
 ID HV3A\_HUMAN  
 AC P01762;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Ig heavy chain V-III region TR0.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE (MYELOMA PROTEIN TR0).  
 RX MEDLINE=76023781; PubMed=809331;  
 RA Klatzin H., Altevogt P., Ruban E., Kortt A., Starosck K.,  
 RA Hilschmann N.;  
 RT "The primary structure of a monoclonal IgA-immunoglobulin (IgA Tr0),  
 RT structure of the complete IgA-molecule."  
 RT Hoppe-Sejler S.Z. Physiol. Chem. 356:1337-1342(1975).  
 CC -1- MISCELLANEOUS: THE SEQUENCE OF THE C REGION IS ALSO GIVEN.  
 CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.  
 DR PIR; A02045; A1HTR.  
 DR HSSP; P01772; 2F84.  
 DR GO; GO:0005576; Cxcracellular; NAS.  
 DR GO; GO:0003823; Piantigen binding; NAS.  
 DR GO; GO:0006955; P:immune response; NAS.  
 DR InterPro; IPR007110; Ig-like.  
 DR InterPro; IPR003596; Ig\_v.  
 DR Pfam; PF00047; Ig; 1.  
 DR Pfam; PF00047; Ig; 1.

DR SMART; SM00406; IGV; 1.  
 DR PROSITE; PSS0835; IG LIKE; 1.  
 KM Immunoglobulin V region; Pyrolydione carboxylic acid.  
 FT DOMAIN 1 108 IG-LIKE.  
 FT MOD\_RES 1 1 PYRROLIDONE CARBOXYLIC ACID.  
 FT NON\_TER 122 122  
 SQ SEQUENCE 122 AA; 13472 MW; 2E21A11DA04D80F9 CRC64;  
 Query Match 45.8%; Score 318.5; DB 1; Length 122;  
 Best Local Similarity 53.3%; Pred. No. 5.9e-24;  
 Matches 65; Conservative 18; Mismatches 38; Indels 1; Gaps 1;  
 QY 1 QVQLDPSGGGLVQVGTGSLRLSCAASGPGIDVYIMGMYRQAPKQREMVASITATGPP-NY 59  
 DB 1 QVQLVDSGGGLVQVGTGSLRLSCAASGPGIDVYIMGMYRQAPKQREMVASITATGPP-NY 60  
 QY 60 TDSVKGKFTISRNDKNTETYLQWNLKPEPTAVYYCAQVRFSSDYTNWGQGTQVTV 119  
 DB 61 ADSVKGKFTISRNDKNTETYLQWNLKPEPTAVYYCAQVRFSSDYTNWGQGTQVTV 120  
 QY 120 SS 121  
 DB 121 SS 122  
 RESULT 13  
 HV33\_MOUSE STANDARD; PRT; 115 AA.  
 ID HV33\_MOUSE  
 AC P01802;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Ig heavy chain V-III region W3082.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE.  
 RX MEDLINE=82099361; PubMed=6798111;  
 RA Johnson N., Stankard J., Paul L.;  
 RT "The complete V domain amino acid sequences of two myeloma immunin-  
 RT binding proteins."  
 RL J. Immunol. 128:302-307(1982).  
 CC -1- MISCELLANEOUS: THIS CHAIN WAS ISOLATED FROM A MYELOMA PROTEIN THAT  
 CC BINDS INULIN.  
 CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.  
 DR PIR; D92811; AYMS82.  
 DR HSSP; P01810; 2F84.  
 DR InterPro; IPR007110; Ig-like.  
 DR InterPro; IPR003596; Ig\_v.  
 DR Pfam; PF00047; Ig; 1.  
 DR SMART; SM00406; IGV; 1.  
 DR PROSITE; PSS0835; IG LIKE; 1.  
 KM Immunoglobulin V region.  
 FT DOMAIN 1 114 IG-LIKE.  
 FT DISULFID 22 98 BY SIMILARITY.  
 FT NON\_TER 115 115  
 SQ SEQUENCE 115 AA; 12887 MW; 9B4517648C121C5A CRC64;  
 Query Match 45.6%; Score 317; DB 1; Length 115;  
 Best Local Similarity 52.4%; Pred. No. 7.7e-24;  
 Matches 65; Conservative 16; Mismatches 31; Indels 12; Gaps 2;  
 QY 1 QVQLDPSGGGLVQVGTGSLRLSCAASGPGIDVYIMGMYRQAPKQREMVASITATG--PP 57  
 DB 1 EVTLDSGGGLVQVGTGSLRLSCAASGPGIDVYIMGMYRQAPKQREMVASITATG 60  
 QY 58 NYTDSVKGKFTISRNDKNTETYLQWNLKPEPTAVYYCAQVRFSSDYTNWGQGTQVTV 117  
 DB 61 HVAESVKGKFTISRNDKNTETYLQWNLKPEPTAVYYCAQVRFSSDYTNWGQGTQVTV 111  
 QY 118 TVSS 121

Db 112 TVSA 115

## RESULT 14

HV3E\_HUMAN STANDARD; PRT; 120 AA.

AC P01766; 21-JUL-1986 (Rel. 01, Created)

DT 21-JUL-1986 (Rel. 01, Last sequence update)

DE 10-OCT-2003 (Rel. 42, Last annotation update)

OS Homo sapiens (Human)

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

NCBI\_TaxId=9606;

RP MEDLINE=77117674; PubMed=65324;

RA Capra J.D., Hopper J.E., Comparative studies on monocytic IgM lambda and IgG kappa from an individual patient. III. The complete amino acid sequence of the VH region of the IgM paraprotein.;

RT Immunohistochemistry 13:995-999(1976).

CC -1- MISCELLANEOUS: THIS CHAIN WAS OBTAINED FROM IGM ISOLATED FROM THE SERUM OF A PATIENT WITH MALIGNANT LYMPHOMA OF THE WALDENSTROM TYPE.

CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.

DR PIR; A02049; M3HUB.

DR HSSP; P01772; 2FB4.

DR GO; GO:0005576; C:extracellular; NAS.

DR GO; GO:0003823; F:antigen binding; NAS.

DR GO; GO:0006955; P:immune response; NAS.

DR InterPro; IPR007110; Ig-like.

DR InterPro; IPR003596; Ig\_v.

DR Pfam; PF00047; Ig\_v.

DR SMART; SM00406; IGV; 1.

DR PROSITE; PS00835; IG LIKE; 1.

DR Immunoglobulin V region.

KW DOVAIN 1 111 IG-LIKE.

FT NON TER 120 120

SQ SEQUENCE 120 AA; 13227 MW; D3F0428F7C2E6410 CRC64;

Query Match 45.6%; Score 317; DB 1; Length 120;

Best Local Similarity 57.5%; Pred. No. 8.1e-24;

Matches 69; Conservative 10; Mismatches 35; Indels 6; Gaps 2;

QY 1 QVQLDSSGGGLVQPGGSLRLSCAASGPGIGDYYLMGWYRQAFGKOREMVASITATGPPNYT 60

DB 1 EVQLVESGGGLVQPGGSLRLSCAASGPGFSYNNMWVQVTKGLEWVSAIGTAGDQYTA 60

QY 61 DSVKGRFTISRDNKNTLEYLQNNLKPEDTAVYYCNQAVRVFSSDYTNV-----WGCGT 115

DB 61 DSVKGRFTISRDNKNTLEYLQNNLKPEDTAVYYCNQAVRVFSSDYTNV-----WGCGT 119

## RESULT 15

HV3H\_HUMAN STANDARD; PRT; 122 AA.

AC P01769; 21-JUL-1986 (Rel. 01, Created)

DT 21-JUL-1986 (Rel. 01, Last sequence update)

DE 10-OCT-2003 (Rel. 42, Last annotation update)

OS Homo sapiens (Human)

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

NCBI\_TaxId=9606;

RP MEDLINE=74175307; PubMed=4208843;

RA Florent G., Lehman D., Putnam F.W.;

RT "The switch point in mu heavy chains of human IgM immunoglobulins.;"

RL Biochemistry 13:2482-2498(1974).

CC -1- MISCELLANEOUS: THIS CHAIN WAS ISOLATED FROM A WALDENSTROM'S MACROGLOBULIN.

CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.

DR PIR; A02052; M3HUG.

DR HSSP; P01772; 2FB4.

DR GO; GO:0005576; C:extracellular; NAS.

DR GO; GO:0003823; F:antigen binding; NAS.

DR GO; GO:0006955; P:immune response; NAS.

DR InterPro; IPR007110; Ig-like.

DR Pfam; PF00047; Ig\_v.

DR SMART; SM00406; IGV; 1.

DR PROSITE; PS00835; IG LIKE; 1.

DR Immunoglobulin V region; Pyroliidone carboxylic acid.

KW DOVAIN 1 112 IG-LIKE.

FT MOD RES 1 122 PYROLIIDONE CARBOXYLIC ACID.

FT NON TER 122 122

SQ SEQUENCE 122 AA; 13166 MW; 745B56959E84100A CRC64;

Query Match 45.5%; Score 316.5; DB 1; Length 122;

Best Local Similarity 54.1%; Pred. No. 9.2e-24;

Matches 66; Conservative 17; Mismatches 38; Indels 1; Gaps 1;

QY 1 QVQLDSSGGGLVQPGGSLRLSCAASGPGIGDYYLMGWYRQAFGKOREMVASITATGPPN-Y 59

DB 1 QVQLVESGGGLVQPGGSLRLSCAASGPGFSYNNMWVQVTKGLEWVSAIGTAGDQYTA 60

QY 60 TDSVKGKFTISRDNKNTLEYLQNNLKPEDTAVYYCNQAVRVFSSDYTNVWGCGTQVTV 119

DB 61 AASVKGKFTISRDBSKBTWYLENNLSRAENTAVYYCARSGIALGSAVAGTDYWGZGLVTVI 120

QY 120 SS 121

DB 121 SS 122

Search completed: October 5, 2004, 08:01:56

Job time : 11.6358 secs

This Page Blank (uspofo)



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2004, 07:34:11 ; Search time 68.817 Seconds

(without alignments)  
596.081 Million cell updates/sec

Title: US-09-805-290A-24

Perfect score: 695  
Sequence: 1 QVQLDSGGGLVOTGSLRL.....WGQCTGVTSSEPKPKPKP 130

Scoring table:  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

1: SP\_ARCHAEA:\*  
2: SP\_BACTERIA:\*  
3: SP\_FUNGI:\*  
4: SP\_HUMAN:\*  
5: SP\_INVERTEBRATE:\*  
6: SP\_MAMMAL:\*  
7: SP\_MHC:\*  
8: SP\_ORGANELLE:\*  
9: SP\_PHAGE:\*  
10: SP\_PLANT:\*  
11: SP RODENT:\*  
12: SP\_VIRUS:\*  
13: SP\_VERTEBRATE:\*  
14: SP\_UNCLASSIFIED:\*  
15: SP\_VIRUS:\*  
16: SP\_BACTERIAD:\*  
17: SP\_ARCHAEA\*:

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	365	52.5	597	4	Q96BB9
2	360.5	51.9	118	4	Q9UL72
3	357	51.4	113	4	Q9UL90
4	352	50.6	573	4	Q8WU38
5	351	50.5	121	4	Q9UL71
6	351	50.5	494	4	Q96K68
7	342.5	49.3	499	4	Q8N5K4
8	342.5	49.3	613	4	Q8WUK1
9	341.5	49.0	437	11	Q9R1A4
10	340.5	49.0	118	4	Q9UL91
11	339.5	48.8	122	4	Q9UL84
12	338.5	48.7	112	4	Q9HCCI
13	337	48.5	116	4	Q9UL93
14	334.5	48.1	487	11	Q99KX4
15	332	47.8	479	11	Q91WP5
16	330	47.5	486	11	Q91Z07

17	329	47.3	119	11	Q92OE7	Q92OE7 mus musculu
18	327	47.1	131	4	Q91B88	Q91B88 homo sapien
19	326	46.9	147	4	Q9Y509	Q9Y509 homo sapien
20	324.5	46.7	473	11	Q91Z05	Q91Z05 mus musculu
21	324	46.6	480	11	Q91XEL	Q91XEL mus musculu
22	322	46.3	487	11	Q80Z17	Q80Z17 mus musculu
23	321.5	46.3	298	11	Q9GYF0	Q9GYF0 mus musculu
24	318.5	45.8	469	11	Q8R3V9	Q8R3V9 mus musculu
25	317.5	45.7	493	4	Q8NCL6	Q8NCL6 homo sapien
26	316.5	45.5	95	4	Q9ULB6	Q9ULB6 homo sapien
27	309.5	44.5	484	11	Q8VEAO	Q8VEAO mus musculu
28	307.5	44.2	521	4	Q8N4Y9	Q8N4Y9 homo sapien
29	300.5	43.2	479	11	Q7YMK4	Q7YMK4 mus musculu
30	299.5	43.1	470	4	Q7ZSW1	Q7ZSW1 homo sapien
31	281	40.4	112	4	Q9UGP3	Q9UGP3 homo sapien
32	277	39.9	104	4	Q9UL87	Q9UL87 homo sapien
33	276.5	39.8	241	11	Q921A6	Q921A6 mus musculu
34	268.5	38.6	124	6	Q9N0W6	Q9N0W6 oryctolagus
35	266.5	38.3	124	6	Q9N0W4	Q9N0W4 oryctolagus
36	265.5	38.2	482	4	Q7Z351	Q7Z351 homo sapien
37	263.5	37.9	482	11	Q91X92	Q91X92 mus musculu
38	258.5	37.2	143	11	Q924R0	Q924R0 mus musculu
39	257.5	37.1	482	11	Q8K172	Q8K172 mus musculu
40	256.5	36.9	488	11	Q91WR1	Q91WR1 mus musculu
41	256	36.8	142	11	Q924Q1	Q924Q1 mus musculu
42	255.5	36.8	145	11	Q924R4	Q924R4 mus musculu
43	255	36.7	119	4	Q9UL73	Q9UL73 homo sapien
44	255	36.7	159	4	Q96QSO	Q96QSO homo sapien
45	252	36.3	117	11	Q90XFO	Q90XFO mus musculu

## ALIGNMENTS

## RESULT 1

Q96BB9 ID Q96BB9 PRELIMINARY; PRT; 597 AA.

AC 01-DBC-2001 (TrEMBLrel. 19, Created)  
DT 01-DBC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Hypothetical protein.  
OS Homo sapiens (human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAUBERG R.;  
RA Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
RL EMBL; BC015760; AAH15760.1; -  
DR InterPro; IPR007110; IG\_1like.  
DR InterPro; IPR003006; IG\_MHC.  
DR InterPro; IPR003596; IG\_V.  
DR Pfam; PF00047; IG\_5.  
DR SMART; SM00406; IGV\_1.  
DR PROSITE; PS00835; IGV\_LIKE; 5.  
DR PROSITE; PS00290; IGV\_MHC; 3.  
KM Hypothetical protein.  
SQ SEQUENCE 597 AA; 65039 MW; 4FCA3AD8CE263D9 CRC64;

Query Match 52.5%; Score 365; DB 4; Length 597;

Best Local Similarity 63.2%; Pred. No. 8.8e-30; Mismatches 30; Indels 4; Gaps 3;

Matches 79; Conservative 12; Mismatches 30; Indels 4; Gaps 3;

QY 1 QVQLDSGGGLVOTGSLRLSCAASGFLDVLWGMYRQAPGKOREMVASITATPPN-Y 59  
Db 20 EVQLLESGGGLVQPGGSLRLSCAASGFSFSYAMVMVRQAPGLEWVAISGSGSTY 79  
QY 60 TDSVKRFTISNDNNTFYLMNNIKPEDTAVYVCNAQVR-VRRSSDT-NYGGQGTQ 116  
Db 80 ADSVKRFTISNDNNTFYLMNNISLRADTAVYVCADKPRYSASAGNTRDYGQGT 139

OY 117 VTSS 121  
 DB 140 VTSS 144

## RESULT 2

O9UL72 PRELIMINARY; PRT; 118 AA.  
 AC O9UL72;  
 DT 01-MAY-2000 (TRENBLrel. 13, Created)  
 DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)  
 DE Myosin-reactive immunoglobulin heavy chain variable region (fragment).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 NCBI\_TaxID=9606;  
 RX SEQUENCE FROM N.A.  
 MEDLINE=98277139; PubMed=9614934;  
 RA Wu X., Liu B., Van der Merwe P.L., Kalis N.N., Berney S.M.,  
 RT "Myosin-reactive autoantibodies in rheumatic carditis and normal fetus."  
 RL Clin. Immunol. Immunopathol. 87:184-192(1998).  
 DR EMBL; AF035042; AAD56278.1; -.  
 DR PIR; S21205; S21205.  
 DR HSSP; P01772; 2FB4.  
 DR InterPro; IPR007110; IG-like.  
 DR InterPro; IPR003596; IG\_V.  
 DR Pfam; PF00047; IG\_1.  
 DR SMART; SM00406; IGV\_1.  
 DR PROSITE; PSS0835; IG-LIKE; 1.  
 FT NON TER 1  
 FT NON TER 118  
 SQ SEQUENCE 118 AA; 12872 MW; B4D1A5944B2DSCCA CRC64;

Query Match 51.9%; Score 360.5; DB 4; Length 118;  
 Best Local Similarity 63.6%; Pred. No. 2.9e-30;

Matches 77; Conservative 7; Mismatches 34; Indels 3; Gaps 1;

OY 1 QVOLDGGGLVOTGSLRLSCAASGPIGDIVYIMGYRQAPGKOREMVASITATGPPNYT 60  
 DB 1 EVOLVSGGGLVOPGSSLRSLSCAASGFTVSNVMWVROAPGKGLSESVYSSGSSYYA 60  
 OY 61 DSVKGRFTISRNDKNTLEYLQMNLIKPEDTAVYYCNAQVRVRFSSDYTNMGQGTQVTVS 120  
 DB 61 DSVKGRFTISRNDKNTLEYLQMNLIKPEDTAVYYCNAQVRVRFSSDYTNMGQGTQVTVS 117  
 OY 121 S 121  
 DB 118 S 118

## RESULT 3

O9UL90 PRELIMINARY; PRT; 113 AA.  
 AC O9UL90;  
 DT 01-MAY-2000 (TRENBLrel. 13, Created)  
 DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)  
 DE Myosin-reactive immunoglobulin heavy chain variable region (fragment).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 NCBI\_TaxID=9606;  
 RX SEQUENCE FROM N.A.  
 MEDLINE=98277139; PubMed=9614934;  
 RA Wu X., Liu B., Van der Merwe P.L., Kalis N.N., Berney S.M.,

RA Young D.C.;  
 RT "Myosin-reactive autoantibodies in rheumatic carditis and normal fetus."  
 RL Clin. Immunol. Immunopathol. 87:184-192(1998).  
 DR EMBL; AF035024; AAD56260.1; -.  
 DR HSSP; P01772; 2FB4.  
 DR InterPro; IPR007110; IG-like.  
 DR InterPro; IPR003596; IG\_V.  
 DR Pfam; PF00047; IG\_1.  
 DR SMART; SM00406; IGV\_1.  
 DR PROSITE; PSS0835; IG-LIKE; 1.  
 FT NON TER 1  
 FT NON TER 113  
 SQ SEQUENCE 113 AA; 12437 MW; ED57FDD19086D07F CRC64;

Query Match 51.4%; Score 357; DB 4; Length 113;  
 Best Local Similarity 63.9%; Pred. No. 6.5e-30;

Matches 78; Conservative 6; Mismatches 28; Indels 10; Gaps 2;

OY 1 QVOLDGGGLVOTGSLRLSCAASGPIGDIVYIMGYRQAPGKOREMVASITATGPPN-Y 59  
 DB 1 EVOLVSGGGLVOPGSSLRSLSCAASGFTVSNVMWVROAPGKGLSESVYSSGSSNYX 60  
 OY 60 TDSVGRFTISRNDKNTLEYLQMNLIKPEDTAVYYCNAQVRVRFSSDYTNMGQGTQVTV 119  
 DB 61 ADSVGRFTISRNDKNTLEYLQMNLIKPEDTAVYYCAKDL-----NYMGQGLTVTV 111  
 OY 120 SS 121  
 DB 112 SS 113

## RESULT 4

O8WU38 PRELIMINARY; PRT; 573 AA.  
 AC O8WU38;  
 DT 01-MAR-2002 (TRENBLrel. 20, Created)  
 DT 01-MAR-2002 (TRENBLrel. 20, Last sequence update)  
 DE Hypothetical protein.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 NCBI\_TaxID=9606;  
 RX SEQUENCE FROM N.A.  
 RP Straussberg R.;  
 RC TISSUE=Tonsil;  
 RA Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC021276; AA821276.1; -.  
 DR PIR; S21205; S21205.  
 DR InterPro; IPR007110; IG-like.  
 DR InterPro; IPR003006; IG\_MHC.  
 DR InterPro; IPR003596; IG\_V.  
 DR Pfam; PF00047; IG\_4.  
 DR SMART; SM00406; IGV\_1.  
 DR PROSITE; PSS0835; IG-LIKE; 4.  
 DR PROSITE; PS00290; IG\_MHC; 2.  
 KW Hypothetical protein.  
 SQ SEQUENCE 573 AA; 62967 MW; FD072344033AC530 CRC64;

Query Match 50.6%; Score 352; DB 4; Length 573;  
 Best Local Similarity 58.7%; Pred. No. 2e-28;  
 Matches 81; Conservative 9; Mismatches 36; Indels 12; Gaps 4;

OY 1 QVOLDGGGLVOTGSLRLSCAASGPIGDIVYIMGYRQAPGKOREMVASIT-ATGPPNY 59  
 DB 20 EVOLVSGGGLVOPGSSLRSLSCAASGFTVSNVMWVROAPGKGLSESVYSSGSSNYX 79  
 OY 60 TDSVGRFTISRNDKNTLEYLQMNLIKPEDTAVYYCNAQVRVRFSSDYTNY-----WGQ 113  
 DB 80 ADSVGRFTISRNDKNTLEYLQMNLIKPEDTAVYYCNAQVRVRFSSDYTNY-----AKHGSSTIGYIGMDWQO 135

XX 19-SEP-2001.  
 PD 20-FEB-2001; 2001EP-00200703.  
 PF 14-MAR-2000; 2000EP-00200930.  
 XX  
 XX (UNIT ) UNILEVER NV.  
 XX (UNIT ) UNILEVER PLC.  
 PA  
 PA Bezemer S, Van De Burg M, De Haard JWM, Tareilus E;  
 PI WPI; 2001-572718/65.  
 DR  
 XX New antibody or its fragments for inhibiting human dietary enzymes,  
 PT useful for cosmetic control of body weight of human beings, comprises  
 PT heavy chain variable domain derived from immunoglobulin naturally devoid  
 PT of light chains.  
 PS  
 XX Example 2; Page 9; 37pp; English.  
 PS The patent discloses antibodies or their fragments comprising a heavy  
 CC chain variable domain (VH) derived from an immunoglobulin naturally  
 CC devoid of light chains specific for inhibiting human dietary enzymes. The  
 CC antibodies or the invention are useful for the preparation of medicaments  
 CC or food for inhibiting the activity of one or more human dietary enzymes  
 CC especially human pancreatic lipase (HPL) or human gastric lipase (HGL)  
 CC which are useful for the cosmetic control of body weight of human beings.  
 CC The present peptide sequence is HPL inhibiting VH fragment, HPL #13 from  
 CC llama (camelid) species  
 SQ Sequence 129 AA;  
 Query Match 72.7%; Score 505.5; DB 4; Length 129;  
 Best Local Similarity 74.6%; Pred. No. 5.9e-34;  
 Matches 97; Conservative 8; Mismatches 24; Indels 1; Gaps 1;  
 QY 1 QVQLDSSGGGVVGTGGSLRLSCAASGIPIDYVMGWYRQAPGRKREMAVSTATGPPNYT 60  
 DB 1 QVQLDSSGGGVVGTGGSLRLSCAASGIPIDYVMGWYRQAPGRKREMAVSTATGPPNYT 60  
 QY 61 DSVKGRFTISRNDKNTLEYLQMNNLKPEDTAVYYCAQVAVRPSDYTNWGGCTQVTVS 120  
 DB 61 DSAKGRFTISKNAKATVTLQNNSLKPEDTAVYYCAAL-TRWDKSVNDYWGCGTVTVS 119  
 QY 121 SEPKTPKPOP 130  
 DB 120 SEPKTPKPOP 129  
 RESULT 6  
 AAEL10559  
 ID AAEL10559 standard; peptide; 130 AA.  
 AC AAEL10559;  
 XX  
 XX 10-DEC-2001 (first entry)  
 DT  
 XX HPL inhibiting VH fragment, HPL #30 from llama species.  
 DE  
 XX llama antibody; camelid; anorectic; heavy chain variable domain; VH;  
 KW human dietary enzyme inhibitor; medicament; human pancreatic lipase; HPL;  
 KW food; human gastric lipase; HGL; cosmetic control; body weight.  
 XX  
 OS Lama sp.  
 XX  
 XX Key Location/Qualifiers  
 FH 31..35  
 FT /label= CDR1  
 FT /note= "Complementarity determining region 1"  
 FT 49..64  
 FT /label= CDR2  
 FT /note= "Complementarity determining region 2"

FT Region 98..110  
 FT /label= CDR3  
 FT /note= "Complementarity determining region 3"  
 XX  
 XX EPI134231-AA.  
 XX  
 XX 19-SEP-2001.  
 PD  
 XX 20-FEB-2001; 2001EP-00200703.  
 PF 14-MAR-2000; 2000EP-00200930.  
 XX  
 XX (UNIT ) UNILEVER NV.  
 XX (UNIT ) UNILEVER PLC.  
 PA  
 PA Bezemer S, Van De Burg M, De Haard JWM, Tareilus E;  
 PI WPI; 2001-572718/65.  
 DR  
 XX New antibody or its fragments for inhibiting human dietary enzymes,  
 PT useful for cosmetic control of body weight of human beings, comprises  
 PT heavy chain variable domain derived from immunoglobulin naturally devoid  
 PT of light chains.  
 PS  
 XX Example 2; Page 10; 37pp; English.  
 PS The patent discloses antibodies or their fragments comprising a heavy  
 CC chain variable domain (VH) derived from an immunoglobulin naturally  
 CC devoid of light chains specific for inhibiting human dietary enzymes. The  
 CC antibodies or the invention are useful for the preparation of medicaments  
 CC or food for inhibiting the activity of one or more human dietary enzymes  
 CC especially human pancreatic lipase (HPL) or human gastric lipase (HGL)  
 CC which are useful for the cosmetic control of body weight of human beings.  
 CC The present peptide sequence is HPL inhibiting VH fragment, HPL #30 from  
 CC llama (camelid) species  
 SQ Sequence 130 AA;  
 Query Match 72.2%; Score 502; DB 4; Length 130;  
 Best Local Similarity 74.6%; Pred. No. 1.2e-33;  
 Matches 97; Conservative 8; Mismatches 25; Indels 0; Gaps 0;  
 QY 1 QVQLDSSGGGVVGTGGSLRLSCAASGIPIDYVMGWYRQAPGRKREMAVSTATGPPNYT 60  
 DB 1 QVQLDSSGGGVVGTGGSLRLSCAASGIPIDYVMGWYRQAPGRKREMAVSTATGPPNYT 60  
 QY 61 DSVKGRFTISRNDKNTLEYLQMNNLKPEDTAVYYCAQVAVRPSDYTNWGGCTQVTVS 120  
 DB 61 DSVKGRFTISRNDKNTLEYLQMNNLKPEDTAVYYCAQVAVRPSDYTNWGGCTQVTVS 120  
 QY 121 SEPKTPKPOP 130  
 DB 121 SEPKTPKPOP 130  
 RESULT 7  
 AAEL10554  
 ID AAEL10554 standard; peptide; 130 AA.  
 AC AAEL10554;  
 XX  
 XX 10-DEC-2001 (first entry)  
 DT  
 XX HPL inhibiting VH fragment, HPL #14 from llama species.  
 DE  
 XX llama antibody; camelid; anorectic; heavy chain variable domain; VH;  
 KW human dietary enzyme inhibitor; medicament; human pancreatic lipase; HPL;  
 KW food; human gastric lipase; HGL; cosmetic control; body weight.  
 XX  
 OS Lama sp.  
 XX  
 XX Key Location/Qualifiers  
 FH 31..35  
 FT /label= CDR1  
 FT /note= "Complementarity determining region 1"  
 FT 49..64  
 FT /label= CDR2  
 FT /note= "Complementarity determining region 2"

PI Bezemer S, Van De Burg M, De Haard JWM, Tareilus E;  
XX WPI, 2001-572718/65.

PT New antibody or its fragments for inhibiting human dietary enzymes,  
XX useful for cosmetic control of body weight of human beings, comprises  
PT heavy chain variable domain derived from immunoglobulin naturally devoid  
PT of light chains.

PS Example 2; Page 9; 37pp; English.

CC The patent discloses antibodies or their fragments comprising a heavy  
CC chain variable domain (VH) derived from an immunoglobulin naturally  
CC devoid of light chains specific for inhibiting human dietary enzymes. The  
CC antibodies of the invention are useful for the preparation of medicaments  
CC or food for inhibiting the activity of one or more human dietary enzymes  
CC especially human pancreatic lipase (HPL) or human gastric lipase (HGL)  
CC which are useful for the cosmetic control of body weight of human beings.  
CC The present peptide sequence is HPL inhibiting VH fragment, HPL #15 from  
CC llama (camelid) species

CC Sequence 131 AA;

Query Match 75.3%; Score 523.5; DB 4; Length 131;  
Best Local Similarity 76.3%; Pred. No. 2e-35; Mismatches 1; Gaps 1;  
Matches 100; Conservative 10; Indels 1; Gaps 1;

QY 1 QVQLDSSGGGLVQTGSLRLSCAASGPTIDYLVNGWYRQAPKQREMYASITATGPENYT 60

DB 1 QVQLDSSGGGLVQTGSLRLSCAASGPTIDYLVNGWYRQAPKQREMYASITATGPENYT 60

QY 61 DSVKGRFTISRDNDKNTIYLQNNLKPEPTAVYYCAQVRFSSDPT-NYWGCGTQVTV 119

DB 61 DSVKGRFTISRDNDKNTIYLQNNLKPEPTAVYYCAQVRFSSDPT-NYWGCGTQVTV 120

QY 120 SEPKTPKPOP 130

DB 121 SEPKTPKPOP 131

RESULT 4

ID AAE10552 standard; peptide; 130 AA.

XX AAE10552;

DT 10-DEC-2001 (first entry)

DE HPL inhibiting VH fragment, HPL #12 from llama species.

XX llama antibody; camelid; anorectic; heavy chain variable domain; VH;

KW human dietary enzyme inhibitor; medicament; human pancreatic lipase; HPL;

XX food; human gastric lipase; HGL; cosmetic control; body weight.

OS Lama sp.

XX Key Location/Qualifiers

FT 31..35 /label= CDR1

FT /note= "Complementarity determining region 1"

FT 50..64 /label= CDR2

FT /note= "Complementarity determining region 2"

FT 98..110 /label= CDR3

FT /note= "Complementarity determining region 3"

XX EP1134231-A1.

XX 19-SEP-2001.

XX 20-FEB-2001; 2001EP-00200703.

PR 14-MAR-2000; 2000EP-00200930.  
XX (UNITL) UNILEVER NV.  
PA (UNITL) UNILEVER PLC.

PI Bezemer S, Van De Burg M, De Haard JWM, Tareilus E;  
XX WPI, 2001-572718/65.

PT New antibody or its fragments for inhibiting human dietary enzymes,  
XX useful for cosmetic control of body weight of human beings, comprises  
PT heavy chain variable domain derived from immunoglobulin naturally devoid  
PT of light chains.

PS Example 2; Page 9; 37pp; English.

CC The patent discloses antibodies or their fragments comprising a heavy  
CC chain variable domain (VH) derived from an immunoglobulin naturally  
CC devoid of light chains specific for inhibiting human dietary enzymes. The  
CC antibodies of the invention are useful for the preparation of medicaments  
CC or food for inhibiting the activity of one or more human dietary enzymes  
CC especially human pancreatic lipase (HPL) or human gastric lipase (HGL)  
CC which are useful for the cosmetic control of body weight of human beings.  
CC The present peptide sequence is HPL inhibiting VH fragment, HPL #12 from  
CC llama (camelid) species

CC Sequence 130 AA;

Query Match 73.8%; Score 513; DB 4; Length 130;  
Best Local Similarity 75.4%; Pred. No. 1.4e-34;  
Matches 98; Conservative 9; Mismatches 23; Indels 0; Gaps 0;

QY 1 QVQLDSSGGGLVQTGSLRLSCAASGPTIDYLVNGWYRQAPKQREMYASITATGPENYT 60

DB 1 QVQLDSSGGGLVQTGSLRLSCAASGPTIDYLVNGWYRQAPKQREMYASITATGPENYT 60

QY 61 DSVKGRFTISRDNDKNTIYLQNNLKPEPTAVYYCAQVRFSSDPT-NYWGCGTQVTV 120

DB 61 DSVKGRFTISRDNDKNTIYLQNNLKPEPTAVYYCAQVRFSSDPT-NYWGCGTQVTV 120

QY 121 SEPKTPKPOP 130

DB 121 SEPKTPKPOP 130

RESULT 5

ID AAE10553 standard; peptide; 129 AA.

XX AAE10553;

DT 10-DEC-2001 (first entry)

DE HPL inhibiting VH fragment, HPL #13 from llama species.

XX llama antibody; camelid; anorectic; heavy chain variable domain; VH;

KW human dietary enzyme inhibitor; medicament; human pancreatic lipase; HPL;

XX food; human gastric lipase; HGL; cosmetic control; body weight.

OS Lama sp.

XX Key Location/Qualifiers

FT 31..35 /label= CDR1

FT /note= "Complementarity determining region 1"

FT 50..64 /label= CDR2

FT /note= "Complementarity determining region 2"

FT 98..109 /label= CDR3

FT /note= "Complementarity determining region 3"

XX EP1134231-A1.

XX The patent discloses antibodies or their fragments comprising a heavy  
 CC chain variable domain (VHH) derived from an immunoglobulin naturally  
 CC devoid of light chains specific for inhibiting human dietary enzymes. The  
 CC antibodies of the invention are useful for the preparation of medicaments  
 CC or food for inhibiting the activity of one or more human dietary enzymes  
 CC especially human pancreatic lipase (HPL) or human gastric lipase (HGL)  
 CC which are useful for the cosmetic control of body weight of human beings.  
 CC The present peptide sequence is HPL inhibiting VHH fragment, HPL #19 from  
 CC llama (camelid) species  
 XX Sequence 130 AA;

Query Match 100.0%; Score 695; DB 4; Length 130;  
 Best Local Similarity 100.0%; Pred. No. 1.7e-49;  
 Matches 130; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 QVQLQDSGGGLVQTGGSLRLSCAASGPIGVYLMGWYRQAPGKQREWVASITATGPNTT 60  
 DB 1 QVQLQDSGGGLVQTGGSLRLSCAASGPIGVYLMGWYRQAPGKQREWVASITATGPNTT 60  
 OY 61 DSVKGRFTISRDNKNTYELQMNLPEDTAVYCNQAVRFPSSDYTYWGQGTQVTVS 120  
 DB 61 DSVKGRFTISRDNKNTYELQMNLPEDTAVYCNQAVRFPSSDYTYWGQGTQVTVS 120  
 OY 121 SEPTKPKPOP 130  
 DB 121 SEPTKPKPOP 130

## RESULT 2

AAE10556  
 ID AAE10556 standard; peptide; 130 AA.

XX AAE10556;

DT 10-DEC-2001 (first entry)

DE HPL inhibiting VHH fragment, HPL #18 from llama species.

XX Llama antibody; camelid; anorectic; heavy chain variable domain; VHH;  
 KW human dietary enzyme inhibitor; medicament; human pancreatic lipase; HPL;  
 KW food; human gastric lipase; HGL; cosmetic control; body weight.

OS Lama sp.

XX Key Location/Qualifiers  
 FH Region 31..35  
 FT /label= CDR1  
 FT /note= "Complementarity determining region 1"  
 FT Region 50..64  
 FT /label= CDR2  
 FT /note= "Complementarity determining region 2"  
 FT Region 98..110  
 FT /label= CDR3  
 FT /note= "Complementarity determining region 3"

PN EP1134231-A1.

DD 19-SEP-2001.

XX 20-FEB-2001; 2001EP-00200703.

XX 14-MAR-2000; 2000EP-00200930.

PA (UNIL ) UNILEVER NV.

PA (UNIL ) UNILEVER PLC.

XX Bezener S, Van De Burg M, De Haard JMW, Tareilus E;

XX WPI; 2001-572718/65.

XX New antibody or its fragments for inhibiting human dietary enzymes.

PT useful for cosmetic control of body weight of human beings, comprises  
 PT heavy chain variable domain derived from immunoglobulin naturally devoid  
 XX of light chains.

XX Example 2; Page 10; 37pp; English.

XX The patent discloses antibodies or their fragments comprising a heavy  
 CC chain variable domain (VHH) derived from an immunoglobulin naturally  
 CC devoid of light chains specific for inhibiting human dietary enzymes. The  
 CC antibodies of the invention are useful for the preparation of medicaments  
 CC or food for inhibiting the activity of one or more human dietary enzymes  
 CC especially human pancreatic lipase (HPL) or human gastric lipase (HGL)  
 CC which are useful for the cosmetic control of body weight of human beings.  
 CC The present peptide sequence is HPL inhibiting VHH fragment, HPL #18 from  
 CC llama (camelid) species

XX Sequence 130 AA;

Query Match 81.0%; Score 563; DB 4; Length 130;  
 Best Local Similarity 80.0%; Pred. No. 1.2e-38;  
 Matches 104; Conservative 10; Mismatches 16; Indels 0; Gaps 0;

OY 1 QVQLQDSGGGLVQTGGSLRLSCAASGPIGVYLMGWYRQAPGKQREWVASITATGPNTT 60  
 DB 1 QVQLQDSGGGLVQTGGSLRLSCAASGPIGVYLMGWYRQAPGKQREWVASITATGPNTT 60  
 OY 61 DSVKGRFTISRDNKNTYELQMNLPEDTAVYCNQAVRFPSSDYTYWGQGTQVTVS 120  
 DB 61 DSVKGRFTISRDNKNTYELQMNLPEDTAVYCNQAVRFPSSDYTYWGQGTQVTVS 120  
 OY 121 SEPTKPKPOP 130  
 DB 121 SEPTKPKPOP 130

## RESULT 3

AAE10555  
 ID AAE10555 standard; peptide; 131 AA.

XX AAE10555;

DT 10-DEC-2001 (first entry)

DE HPL inhibiting VHH fragment, HPL #15 from llama species.

XX Llama antibody; camelid; anorectic; heavy chain variable domain; VHH;  
 KW human dietary enzyme inhibitor; medicament; human pancreatic lipase; HPL;  
 KW food; human gastric lipase; HGL; cosmetic control; body weight.

OS Lama sp.

XX Key Location/Qualifiers  
 FH Region 31..35  
 FT /label= CDR1  
 FT /note= "Complementarity determining region 1"  
 FT Region 50..64  
 FT /label= CDR2  
 FT /note= "Complementarity determining region 2"  
 FT Region 98..111  
 FT /label= CDR3  
 FT /note= "Complementarity determining region 3"

PN EP1134231-A1.

DD 19-SEP-2001.

XX 20-FEB-2001; 2001EP-00200703.

XX 14-MAR-2000; 2000EP-00200930.

PA (UNIL ) UNILEVER NV.

PA (UNIL ) UNILEVER PLC.

OM protein - protein search, using sw model

Run on: October 5, 2004, 07:04:10 ; Search time 98.9043 Seconds  
(without alignments)

371.381 Million cell updates/sec

Title: US-09-805-290A-24

Perfect score: 695  
Sequence: 1 QVGIQDSGGGLVQTGGSLR.....WGQGIQVTVSSBPXTPKPP 130

Scoring table: BLOSUM62

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

Post-processing:	Minimum Match 0%
	Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_29Jan04:\*

```
A_Geneseq_29Jan04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysts of the total score distribution.

Result No.	Query Score	Match Length	DB ID	Description
1	100	100	1	...

1	695	100.0	130	4	AAE10557	AAE10557	HPL	inhib
2	563	81.0	130	4	AAE10556	AAE10556	HPL	inhib
3	523.5	75.3	131	4	AAE10555	AAE10555	HPL	inhib
4	513	73.8	130	4	AAE10552	AAE10552	HPL	inhib
5	505.5	72.7	129	4	AAE10553	AAE10553	HPL	inhib
6	502	72.2	130	4	AAE10559	AAE10559	HPL	inhib
7	500	71.9	130	4	AAE10554	AAE10554	HPL	inhib
8	492	70.8	124	4	AAE10563	AAE10563	HGL	inhib
9	487.5	70.1	129	4	AAE10556	AAE10556	HPL	inhib
10	487	70.1	124	4	AAE10565	AAE10565	HGL	inhib
11	476	68.3	130	4	AAE10567	AAE10567	HGL	inhib
12	475.5	68.4	131	5	ABG30620	ABG30620	Immuno	glo
13	472	67.9	152	4	AAE05283	AAE05283	Anti-pota	
14	467.5	67.3	129	4	AAE10561	AAE10561	HPL	inhib
15	465.5	67.0	149	4	AAE05288	AAE05288	Anti-pota	
16	463.5	66.7	129	4	AAE10551	AAE10551	HPL	inhib
17	463.5	66.7	225	2	AAV41165	AAV41165	Llama Vhh	
18	460.5	66.3	129	4	AAE10564	AAE10564	HGL	inhib
19	455	66.0	128	4	AAE10565	AAE10565	HGL	inhib
20	457	65.8	130	7	ABR62883	ABR62883	Llama ant	
21	456	65.6	130	7	ABR62882	ABR62882	Llama ant	
22	455.5	65.5	133	5	ABG30572	ABG30572	Immuno	glo
23	454	65.3	134	5	ABG30618	ABG30618	Immuno	glo
24	453	65.2	190	7	ABR62881	ABR62881	Llama ant	
25	452	65.0	130	7	ABR62879	ABR62879	Llama ant	

26	451	64.9	177	ABR62880	Llama ant
27	445.5	64.1	190	ABR50586	ABR62880 Llama ant
28	445.5	64.1	383	AAE05287	AAE05286 Llama ant
29	441.5	63.5	124	AAE28615	AAE05287 Llama ant
30	441	63.5	119	ABR67777	ABR67776 Llama ant
31	439.5	63.2	204	AAE41167	ABR67777 Antro aci
32	439.5	63.2	205	AAE41173	AAE41167 Llama ant
33	437	62.9	119	ABR67778	AAE41173 Llama ant
34	437	62.9	208	ABR13562	ABR67778 Antro aci
35	436	62.7	2	AAE41168	AAE10562 HGL inhib
36	435	62.7	4	ABR67779	AAE41168 Llama ant
37	434.5	62.5	119	ABR67779	ABR67779 Antro aci
38	432.5	62.2	127	ABR430617	ABR67779 Mouse
39	429.5	61.8	5	AAE41176	ABR430617 Llama ant
40	429.5	61.8	219	AAE41172	AAE41176 Llama ant
41	424.5	61.0	138	ABR62878	ABR62878 Llama ant
42	423	60.9	194	ABR62885	ABR62885 Llama ant
43	422	60.7	214	AAE41178	ABR62885 Llama ant
44	420	60.4	217	AAE41175	AAE41178 Llama ant
45	418	60.1	111	AAE39816	AAE41175 Llama ant

AAE10557  
ID AAE10557 standard; peptide; 130 AA

AAE10557;

10-DEC-2001 (First entry)

HPL inhibiting VHH fragment, HPL #19 from llama species.

Llama antibody; camelid; anorectic; heavy chain variable domain; VHH; human dietary enzyme inhibitor; medicament; human pancreatic lipase; HPL; food; human gastric lipase; HGL; cosmetic control; body weight.

**Lama sp**

Key	Location/Qualifiers
31	35

/label=  
/note=

Region	50. .64
71 abel	

/notes  
08 71

```

/label=
/notes=

```

FD-311 (Rev. 1-25-60)

2  
3  
4  
5  
6  
7  
8

2001 FEB - 002

14 MAR 2000: 2000FB-003

UNITED STATES OF AMERICA

(UNIL ) UNILEVER PLC.

Bezemer, S., Van De Burgh

WPI; 2001-572718/65.

New antibody or its fra

heavy chain variable domain of light chains

From 10: 30: Page 10: 37

100

Tue Oct 5 09:51:22 2004

us-09-805-290a-24.rsp

Page 6

OY 61 DSVKGRFTISRNDKNTETYLQNNNLKPEDTAVYYCNAQVRVRFSSDYTNMGQGTQVTVS 120  
DB 61 DSVKGRFTISRNDKNTETYLQNNNLKPEDTAVYYCNAQVRVRFSSDYTNMGQGTQVTVS 115  
OY 121 S 121  
DB 116 S 116

RESULT 14

O99KX4 PRELIMINARY; PRT; 487 AA.  
AC O99KX4;  
DT 01-JUN-2001 (Tremblrel. 17, Created)  
DT 01-JUN-2001 (Tremblrel. 17, last sequence update)  
DT 01-OCT-2003 (Tremblrel. 25, last annotation update)  
DE Hypothetical protein.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Strausberg R.;  
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL: BC004786; AAH04786.1; -  
DR HSSP: P01810; 2FBL.  
DR InterPro: IPR007110; Ig-like.  
DR InterPro: IPR003006; Ig\_MHC.  
DR InterPro: IPR003596; Ig\_Y.  
DR Pfam: PF00047; Ig; 4.  
DR SMART: SM00406; Ig; 1.  
DR PROSITE: PS00835; IG\_LIKE; 4.  
DR PROSITE: PS00230; IG\_MHC; 2.  
KW Hypothetical protein.  
SQ SEQUENCE 487 AA; 52554 MW; 7DC8E96DB33077B CRC64;

Query Match 48.1%; Score 334.5; DB 11; Length 487;  
Best Local Similarity 57.8%; Pred. No. 1.1e-26;  
Matches 74; Conservative 13; Mismatches 32; Indels 9; Gaps 3;

OY 1 QVQLDSGGGLVGTGSLRLSCAASGPIGDIYIMGWYRQAPGKQRMVASITATGP-PHY 59  
DB 20 EVQLVESGGGLVPRGSLKLSCAASGPTFSYAMSVRQTPPEKRLVVAALNSNGNTYY 79  
OY 60 TDSVKGRFTISRNDKNTETYLQNNNLKPEDTAVYYCNAQVRVRFSSDY-----TNMGQ 114  
DB 80 PDVVKGRFTISRNDKNTETYLQNNNLKPEDTAVYYCNAQVRVRFSSDY-----TNMGQ 136  
OY 115 TQVTVSSE 122  
DB 137 TTVTVSSE 144

RESULT 15

O91WP5 PRELIMINARY; PRT; 479 AA.  
AC O91WP5;  
DT 01-DEC-2001 (Tremblrel. 19, Created)  
DT 01-DEC-2001 (Tremblrel. 19, last sequence update)  
DT 01-OCT-2003 (Tremblrel. 25, last annotation update)  
DE Hypothetical protein.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Strausberg R.;  
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL: BC013656; AAH13656.1; -

DR InterPro: IPR007110; Ig-like.  
DR InterPro: IPR003006; Ig\_MHC.  
DR InterPro: IPR003596; Ig\_Y.  
DR Pfam: PF00047; Ig; 4.  
DR SMART: SM00406; Ig; 1.  
DR PROSITE: PS00835; IG\_LIKE; 4.  
DR PROSITE: PS00230; IG\_MHC; 2.  
KW Hypothetical protein.  
SQ SEQUENCE 479 AA; 51603 MW; ECB2D087748584F CRC64;

Query Match 47.8%; Score 332; DB 11; Length 479;  
Best Local Similarity 57.3%; Pred. No. 2e-26;  
Matches 71; Conservative 17; Mismatches 28; Indels 9; Gaps 3;

OY 1 QVQLDSGGGLVGTGSLRLSCAASGPIGDIYIMGWYRQAPGKQRMVASITATGP-PHY 59  
DB 20 EVQLVESGGGLVPRGSLKLSCAASGPTFSYAMSVRQTPPEKRLVVAALNSNGNTYY 79  
OY 60 TDSVKGRFTISRNDKNTETYLQNNNLKPEDTAVYYCNAQVRVRFSSDYTNMGQGTQVTV 119  
DB 80 SDTVKGRFTISRNDKNTETYLQNNNLKPEDTAVYYCNAQVRVRFSSDYTNMGQGTQVTV 132  
OY 120 SSEP 123  
DB 133 SSEP 136

Search completed: October 5, 2004, 08:13:48  
Job time : 69.817 secs

Matches 76; Conservative 10; Mismatches 30; Indels 5; Gaps 2;

QY 1 QVOQDGGGGLVOTGSGLRISCAASGPIDGYLMGWYRQAPGKOREMVASITAT-GPPNY 59  
 DB 1 EVQVBSGGGGLVOTGSGLRISCAASGFTFSSYMNWVROAPGKLEWVASITSSITITIT 60  
 QY 60 TDSVKGFTISRDNKNTLEYLQNNMLKPEDTAVVYCNAAQVRVPSSDYTNWGGGTQVY 119  
 DB 61 ADSVKGFTISRDNKNTLEYLQNNMLKPEDTAVVYCNAAQVRVPSSDYTNWGGGTQVY 116

QY 120 S 120  
 DB 117 S 117

## RESULT 11

Q9UL84 PRELIMINARY; PRT; 122 AA.

AC Q9UL84  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
 DE Myosin-reactive immunoglobulin heavy chain variable region (Fragment).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OC NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98277139; PubMed=9614934;  
 RA Wu X., Liu B., Van der Merwe P.L., Kalis N.N., Berner S.N.,  
 RA Young D.C.;  
 RT "Myosin-reactive autoantibodies in rheumatic carditis and normal  
 RT fetus.";  
 RT Clin. Immunol. Immunopathol. 87:184-192(1998).

DR EMBL; P01772; 2FB4.  
 DR HSSP; P01772; 2FB4.  
 DR InterPro; IPR007110; IG-like.  
 DR InterPro; IPR003596; IG\_v.  
 DR Pfam; PF00047; Ig; 1.  
 DR SMART; SM00406; IGV; 1.  
 DR PROSITE; PS50835; IG\_LIKE; 1.  
 DR NON\_TER 1  
 FT NON\_TER 122  
 FT NON\_TER 122  
 SQ SEQUENCE 122 AA; 13579 MW; 36054D41366545B8 CRC64;

Query Match 48.8%; Score 339.5; DB 4; Length 122;  
 Best Local Similarity 61.5%; Pred. No. 5.1e-28;  
 Matches 75; Conservative 8; Mismatches 38; Indels 1; Gaps 1;

QY 1 QVOQDGGGGLVOTGSGLRISCAASGPIDGYLMGWYRQAPGKOREMVASITATGPPN-Y 59  
 DB 1 EVQVBSGGGGLVOTGSGLRISCAASGFTFSSYMNWVROAPGKLEWVASITSSITITIT 60  
 QY 60 TDSVKGFTISRDNKNTLEYLQNNMLKPEDTAVVYCNAAQVRVPSSDYTNWGGGTQVY 119  
 DB 61 ADSVKGFTISRDNKNTLEYLQNNMLKPEDTAVVYCNAAQVRVPSSDYTNWGGGTQVY 120

QY 120 SS 121  
 DB 121 SS 122

## RESULT 12

Q9HCC1 PRELIMINARY; PRT; 112 AA.

AC Q9HCC1  
 DT 01-MAR-2001 (TREMBLrel. 16, Created)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
 DE Single chain Fv (Fragment).  
 OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OC NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Kikuchi M., Ikeda C., Tsujimoto Y., Asada S., Nagata K.;  
 RT "An antibody fragment 2A3 specific for native lysozyme: isolation from a  
 RT human synthetic phage display library and characterization."  
 RL Submitted (OCT-2000) to the EMBL/Genbank/DBJ databases.  
 DR EMBL; AB049915; BAB16829.1; -  
 DR HSSP; P01772; 2FB4.  
 DR InterPro; IPR007110; IG-like.  
 DR InterPro; IPR003596; IG\_v.  
 DR Pfam; PF00047; IGV; 1.  
 DR SMART; SM00406; IGV; 1.  
 DR PROSITE; PS50835; IG\_LIKE; 1.  
 DR NON\_TER 1  
 FT NON\_TER 112  
 FT NON\_TER 112  
 SQ SEQUENCE 112 AA; 12243 MW; 2A1A45EC3B84788 CRC64;

Query Match 48.7%; Score 338.5; DB 4; Length 112;  
 Best Local Similarity 61.9%; Pred. No. 5.8e-28;  
 Matches 73; Conservative 11; Mismatches 27; Indels 7; Gaps 3;

QY 1 QVOQDGGGGLVOTGSGLRISCAASGPIDGYLMGWYRQAPGKOREMVASITAT-GPPNY 59  
 DB 1 EVQVBSGGGGLVOTGSGLRISCAASGFTFSSYMNWVROAPGKLEWVASITSSITITIT 60  
 QY 60 TDSVKGFTISRDNKNTLEYLQNNMLKPEDTAVVYCNAAQVRVPSSDYTNWGGGTQVY 117  
 DB 61 ADSVKGFTISRDNKNTLEYLQNNMLKPEDTAVVYCNAAQVRVPSSDYTNWGGGTQVY 112

## RESULT 13

Q9UL93 PRELIMINARY; PRT; 116 AA.

AC Q9UL93  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
 DE Myosin-reactive immunoglobulin heavy chain variable region (Fragment).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OC NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98277139; PubMed=9614934;  
 RA Wu X., Liu B., Van der Merwe P.L., Kalis N.N., Berner S.N.,  
 RA Young D.C.;  
 RT "Myosin-reactive autoantibodies in rheumatic carditis and normal  
 RT fetus.";  
 RT Clin. Immunol. Immunopathol. 87:184-192(1998).

DR EMBL; AF035021; AAD56257.1; -  
 DR PIR; P01020; P01020.  
 DR HSSP; P01772; 2FB4.  
 DR InterPro; IPR007110; IG-like.  
 DR InterPro; IPR003596; IG\_v.  
 DR Pfam; PF00047; IGV; 1.  
 DR SMART; SM00406; IGV; 1.  
 DR PROSITE; PS50835; IG\_LIKE; 1.  
 DR NON\_TER 1  
 FT NON\_TER 116  
 FT NON\_TER 116  
 SQ SEQUENCE 116 AA; 12434 MW; 0DA0348154D6061 CRC64;

Query Match 48.5%; Score 337; DB 4; Length 116;  
 Best Local Similarity 62.0%; Pred. No. 8.8e-28;  
 Matches 75; Conservative 7; Mismatches 33; Indels 6; Gaps 2;

QY 2 VQLODGGGGLVOTGSGLRISCAASGPIDGYLMGWYRQAPGKOREMVASITATGPPN-YT 60  
 DB 1 EVQVBSGGGGLVOTGSGLRISCAASGFTFSSYMNWVROAPGKLEWVASITSSITITIT 60



QY 60 TDSVKGFTISRDNKTEYLQNNLKPEPTAVYYCAQVRPSSD-----YTNWQ 113  
 DB 80 ADVSKGFTISRDNKTEYLQNNLKPEPTAVYYCAQVRPSSD-----YTNWQ 113  
 QY 114 GTQVTVSS-EPKPKPKP 130  
 DB 139 GTQVTVSSASPTSPKVP 156

RESULT 8

Q8WTK1 ID O8WTK1 PRELIMINARY; PRT; 613 AA.  
 AC O8WTK1  
 DT 01-MAY-2002 (TrEMBLrel. 20, Created)  
 DT 01-MAY-2002 (TrEMBLrel. 20, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Hypothetical protein.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TrEMBLrel. 20, Created  
 RA Strauberg R.  
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: BC020240; AA020240.1; .  
 DR PIR: P0120; P0120.  
 DR PIR: S15590; S15590.  
 DR InterPro: IPR007110; IG\_1like.  
 DR InterPro: IPR003006; IG\_MHC.  
 DR InterPro: IPR003596; IG\_V.  
 DR Pfam: PF00047; IG\_1.  
 DR SMART: SM00406; IG\_1.  
 DR PROSITE: PSS0835; IG\_LIKE; 5.  
 DR PROSITE: PS00290; IG\_MHC; 3.  
 KW Hypothetical protein.  
 SQ SEQUENCE 613 AA; 67296 MW; 60CF75950671E315 CRC64;

Query Match 49.3%; Score 342.5; DB 4; Length 613;  
 Best Local Similarity 60.2%; Pred. No. 2.2e-27;  
 Matches 77; Conservative 9; Mismatches 27; Indels 15; Gaps 3;

QY 1 QVQLODSGGGLVOTGSLRLSCAASGPIGPIVLMGWYRQAPGKOREMVASITATGPPNYTD 61  
 DB 20 QVQLODSGGGLVOTGSLRLSCAASGPIGPIVLMGWYRQAPGKOREMVASITATGPPNYTD 61  
 QY 60 TDSVKGFTISRDNKTEYLQNNLKPEPTAVYYCAQVRPSSD-----YTNWQ 113  
 DB 80 ADVSKGFTISRDNKTEYLQNNLKPEPTAVYYCAQVRPSSD-----YTNWQ 113  
 QY 114 GTQVTVSS 121  
 DB 132 GTQVTVSS 139

RESULT 9

Q9RIA4 ID O9RIA4 PRELIMINARY; PRT; 437 AA.  
 AC O9RIA4  
 DT 01-MAY-2002 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2002 (TrEMBLrel. 13, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Gamma1 heavy chain of Mab7 (Fragment).  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TrEMBLrel. 13, Created  
 RA Wilde K.G., Yu X., Ekramoddoullah A.K.M., Misra S.;

RT "Cloning of cDNAs encoding for anti-white pine blister rust monoclonal  
 RT antibody (Mab 7, its light and heavy chains) and construction of a  
 RT single chain antibody (scFv)."  
 RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF153372; AA040243.1; .  
 DR PIR: B45837; B45837.  
 DR PDB: 1COK; 11-SEP-99.  
 DR PDB: 1COK; 25-DEC-02.  
 DR PDB: 1COK; 11-MAY-02.  
 DR MGI: MGI:96446; 19h-4.  
 DR InterPro: IPR007110; IG\_1like.  
 DR InterPro: IPR003006; IG\_MHC.  
 DR InterPro: IPR003596; IG\_V.  
 DR Pfam: PF00047; IG\_1.  
 DR SMART: SM00406; IG\_1.  
 DR PROSITE: PSS0835; IG\_LIKE; 4.  
 DR PROSITE: PS00290; IG\_MHC; 1.  
 FT NON\_TER 1  
 FT NON\_TER 1  
 SQ SEQUENCE 437 AA; 48142 MW; 5C3A7BB3EE7D697C CRC64;

Query Match 49.1%; Score 341.5; DB 11; Length 437;  
 Best Local Similarity 56.8%; Pred. No. 1.8e-27;  
 Matches 71; Conservative 14; Mismatches 33; Indels 7; Gaps 1;

QY 2 VOLQDSGGGLVOTGSLRLSCAASGPIGPIVLMGWYRQAPGKOREMVASITATGPPNYTD 61  
 DB 1 VOLQDSGGGLVOTGSLRLSCAASGPIGPIVLMGWYRQAPGKOREMVASITATGPPNYTD 61  
 QY 62 SVKGRFTISRDNKTEYLQNNLKPEPTAVYYCAQVRPSSD-----YTNWQ 121  
 DB 61 SVKGRFTISRDNKTEYLQNNLKPEPTAVYYCAQVRPSSD-----YTNWQ 121  
 QY 122 EPKTP 126  
 DB 114 AKTTP 118

RESULT 10

Q9UL91 ID Q9UL91 PRELIMINARY; PRT; 118 AA.  
 AC Q9UL91  
 DT 01-MAY-2002 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2002 (TrEMBLrel. 13, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Myosin-reactive immunoglobulin heavy chain variable region  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TrEMBLrel. 13, Created  
 RA MEDLINE:98277139; PubMed:9614934;  
 RA Wu X., Liu B., Van der Merwe P.L., Kalis N.N., Berny S.M.,  
 RA Young D.C.;  
 RT "Myosin-reactive autoantibodies in rheumatic carditis and normal  
 RT fetus";  
 RT Clin. Immunol. Immunopathol. 87:184-192(1998).  
 DR EMBL: AF035023; AA056259.1; .  
 DR PIR: S21205; S21205.  
 DR HSR: P01772; 2F84.  
 DR InterPro: IPR007110; IG\_1like.  
 DR InterPro: IPR003596; IG\_V.  
 DR Pfam: PF00047; IG\_1.  
 DR SMART: SM00406; IG\_1.  
 DR PROSITE: PSS0835; IG\_LIKE; 1.  
 DR PROSITE: PS00290; IG\_MHC; 1.  
 FT NON\_TER 1  
 FT NON\_TER 1  
 SQ SEQUENCE 118 AA; 12843 MW; D0633949F2AC149D CRC64;

Query Match 49.0%; Score 340.5; DB 4; Length 118;  
 Best Local Similarity 62.8%; Pred. No. 3.8e-28;



FT /label= CDR1  
 FT /note= "Complementarity determining region 1"  
 FT 50..64  
 FT /label= CDR2  
 FT /note= "Complementarity determining region 2"  
 FT 98..110  
 FT /label= CDR3  
 FT /note= "Complementarity determining region 3"

XX EF1134231-A1.

XX 19-SEP-2001.

XX 20-FEB-2001; 2001EP-00200703.

XX 14-MAR-2000; 2000EP-00200930.

XX (UNIL ) UNILEVER NV.

XX (UNIL ) UNILEVER PLC.

XX Bezemer S, Van De Burg M, De Haard JMW, Tareilus E;

XX WPI; 2001-572718/65.

XX New antibody or its fragments for inhibiting human dietary enzymes,  
 PT useful for cosmetic control of body weight of human beings, comprises  
 PT heavy chain variable domain derived from immunoglobulin naturally devoid  
 PT of light chains.

XX Example 2; Page 9; 37pp; English.

XX The patent discloses antibodies or their fragments comprising a heavy  
 CC chain variable domain (VHH) derived from an immunoglobulin naturally  
 CC devoid of light chains specific for inhibiting human dietary enzymes. The  
 CC antibodies of the invention are useful for the preparation of medicaments  
 CC or food for inhibiting the activity of one or more human dietary enzymes  
 CC especially human pancreatic lipase (HPL) or human gastric lipase (HGL)  
 CC which are useful for the cosmetic control of body weight of human beings.  
 CC The present peptide sequence is HPL inhibiting VHH fragment, HPL #14 from  
 CC llama (camelid) species

XX Sequence 130 AA;

Query Match 71.9%; Score 500; DB 4; Length 130;

Best Local Similarity 73.8%; Pred. No. 1.7e-33;

Matches 96; Conservative 10; Mismatches 24; Indels 0; Gaps 0;

QY 1 QVQLDSGGGLVQTGGSLRLSCAASGPIGVLMGWYRQAPGKQREMVASITATGPPNYT 60

Db 1 QVQLDSGGGLVQAGGSLRLSCAASGSGISHTMGWYRQTPGTERDVAVTIQDGGSTNYA 60

QY 61 DSVKGRFTISRDNKNTLEYLQNNLKPEDTAVYVCNAQVRVRFSSDYTNWGGQTQVTVS 120

Db 61 DSVKGRFTISRDNILNTVYLQNNLKPEDTAVVHCNADVPYRTSKYLELWGQGLTVTVS 120

QY 121 SEPKTKPQP 130

Db 121 SEPKTKPQP 130

RESULT 8

AAE10563

ID AAE10563 standard; peptide; 124 AA.

XX AAE10563;

XX 10-DEC-2001 (first entry)

XX HGL inhibiting VHH fragment, HGL #9 from llama species.

XX Llama antibody; camelid; anorectic; heavy chain variable domain; VHH;  
 KW human dietary enzyme inhibitor; medicament; human pancreatic lipase; HPL;  
 KW food; human gastric lipase; HGL; cosmetic control; body weight.

XX Lama sp.  
 OS Key Location/Qualifiers  
 XX 31..35  
 FT /label= CDR1  
 FT /note= "Complementarity determining region 1"  
 FT 50..64  
 FT /label= CDR2  
 FT /note= "Complementarity determining region 2"  
 FT 98..104  
 FT /label= CDR3  
 FT /note= "Complementarity determining region 3"

XX EF1134231-A1.

XX 19-SEP-2001.

XX 20-FEB-2001; 2001EP-00200703.

XX 14-MAR-2000; 2000EP-00200930.

XX (UNIL ) UNILEVER NV.

XX (UNIL ) UNILEVER PLC.

XX Bezemer S, Van De Burg M, De Haard JMW, Tareilus E;

XX WPI; 2001-572718/65.

XX New antibody or its fragments for inhibiting human dietary enzymes,  
 PT useful for cosmetic control of body weight of human beings, comprises  
 PT heavy chain variable domain derived from immunoglobulin naturally devoid  
 PT of light chains.

XX Example 4; Page 13; 37pp; English.

XX The patent discloses antibodies or their fragments comprising a heavy  
 CC chain variable domain (VHH) derived from an immunoglobulin naturally  
 CC devoid of light chains specific for inhibiting human dietary enzymes. The  
 CC antibodies of the invention are useful for the preparation of medicaments  
 CC or food for inhibiting the activity of one or more human dietary enzymes  
 CC especially human pancreatic lipase (HPL) or human gastric lipase (HGL)  
 CC which are useful for the cosmetic control of body weight of human beings.  
 CC The present peptide sequence is HGL inhibiting VHH fragment, HGL #9 from  
 CC llama (camelid) species

XX Sequence 124 AA;

Query Match 70.8%; Score 492; DB 4; Length 124;

Best Local Similarity 73.1%; Pred. No. 7.2e-33;

Matches 95; Conservative 9; Mismatches 20; Indels 6; Gaps 1;

QY 1 QVQLDSGGGLVQTGGSLRLSCAASGPIGVLMGWYRQAPGKQREMVASITATGPPNYT 60

Db 1 QVQLDSGGGLVQAGGSLRLSCAASGSGISLYVMSWYRQAPGKQREPVAAALMGSGSTTYA 60

QY 61 DSVKGRFTISRDNKNTLEYLQNNLKPEDTAVYVCNAQVRVRFSSDYTNWGGQTQVTVS 120

Db 61 DSVKGRFTISRDNKNTMYLQNNLKPEDTGVYYCAG-----TGAEGHYWGQGTQVTVS 114

QY 121 SEPKTKPQP 130

Db 115 SEPKTKPQP 124

RESULT 9

AAE10558

ID AAE10558 standard; peptide; 129 AA.

XX AAE10558;

XX 10-DEC-2001 (first entry)

DE HPL inhibiting VHH fragment, HPL #22 from llama species.  
 XX  
 XX Llama antibody; camelid; anorectic; heavy chain variable domain; VHH;  
 KW human dietary enzyme inhibitor; medicament; human pancreatic lipase; HPL;  
 KW food; human gastric lipase; HGL; cosmetic control; body weight.  
 XX  
 OS Lama sp.  
 XX  
 XX Location/Qualifiers  
 FH Key 31..35  
 FT Region /label= CDR1  
 FT /note= "Complementarity determining region 1"  
 FT Region 50..64  
 FT /label= CDR2  
 FT /note= "Complementarity determining region 2"  
 FT Region 98..109  
 FT /label= CDR3  
 FT /note= "Complementarity determining region 3"  
 XX  
 PN EP1134231-A1.  
 XX  
 XX 19-SEP-2001.  
 XX  
 XX 20-FEB-2001; 2001EP-00200703.  
 XX  
 XX 14-MAR-2000; 2000EP-00200930.  
 XX  
 XX (UNIL ) UNILEVER NV.  
 PA (UNIL ) UNILEVER PLC.  
 XX  
 XX Bezemer S, Van De Burg M, De Haard JJW, Tareilus E;  
 PI WPI; 2001-572718/65.  
 XX  
 DR New antibody or its fragments for inhibiting human dietary enzymes,  
 XX useful for cosmetic control of body weight of human beings, comprises  
 PT heavy chain variable domain derived from immunoglobulin naturally devoid  
 PT of light chains.  
 XX  
 XX Example 2; Page 10; 37pp; English.  
 XX  
 CC The patent discloses antibodies or their fragments comprising a heavy  
 CC chain variable domain (VHH) derived from an immunoglobulin naturally  
 CC devoid of light chains specific for inhibiting human dietary enzymes. The  
 CC antibodies of the invention are useful for the preparation of medicaments  
 CC or food for inhibiting the activity of one or more human dietary enzymes  
 CC especially human pancreatic lipase (HPL) or human gastric lipase (HGL)  
 CC which are useful for the cosmetic control of body weight of human beings.  
 CC The present peptide sequence is HPL inhibiting VHH fragment, HPL #22 from  
 CC llama (camelid) species  
 XX  
 SQ Sequence 129 AA;  
 XX  
 Query Match 70.1%; Score 487.5; DB 4; Length 129;  
 Best Local Similarity 71.5%; Pred. No. 1.8e-32;  
 Matches 93; Conservative 14; Mismatches 22; Indels 1; Gaps 1;  
 QY 1 QVQLDSGGGLVQTGSLRLSCAASGPIGDVYLMGWYRQAPGKOREMVASITATGPPNYT 60  
 DB 1 QVQLDSGGGLVQAGSLRLSCAASGSIRSIIMTYRQAPGKERELVARMSSDGTTSYT 60  
 QY 61 DSVKGRFTISRDNKNTYELQNNLKPEDTAVYYCNQAVRVRFSSDYTNKYGQGTQVTVS 120  
 DB 61 DSVKGRFTISRDNKNTYELQNNLKPEDTAVYYCNQAVRVRFSSDYTNKYGQGTQVTVS 119  
 QY 121 SEPKTPKQP 130  
 DB 120 SEPKTPKQP 129  
 RESULT 10  
 ID AAE10566  
 AAE10566 standard; peptide; 124 AA.

XX AAE10566;  
 AC 10-DEC-2001 (first entry)  
 XX  
 DT HGL inhibiting VHH fragment, HGL #15 from llama species.  
 XX  
 DE Llama antibody; camelid; anorectic; heavy chain variable domain; VHH;  
 KW human dietary enzyme inhibitor; medicament; human pancreatic lipase; HPL;  
 KW food; human gastric lipase; HGL; cosmetic control; body weight.  
 XX  
 OS Lama sp.  
 XX  
 XX Location/Qualifiers  
 FH Key 30..35  
 FT Region /label= CDR1  
 FT /note= "Complementarity determining region 1"  
 FT Region 50..64  
 FT /label= CDR2  
 FT /note= "Complementarity determining region 2"  
 FT Region 98..104  
 FT /label= CDR3  
 FT /note= "Complementarity determining region 3"  
 XX  
 PN EP1134231-A1.  
 XX  
 XX 19-SEP-2001.  
 XX  
 XX 20-FEB-2001; 2001EP-00200703.  
 XX  
 XX 14-MAR-2000; 2000EP-00200930.  
 PR (UNIL ) UNILEVER NV.  
 PA (UNIL ) UNILEVER PLC.  
 XX  
 XX Bezemer S, Van De Burg M, De Haard JJW, Tareilus E;  
 PI WPI; 2001-572718/65.  
 XX  
 DR New antibody or its fragments for inhibiting human dietary enzymes,  
 XX useful for cosmetic control of body weight of human beings, comprises  
 PT heavy chain variable domain derived from immunoglobulin naturally devoid  
 PT of light chains.  
 XX  
 XX Example 4; Page 14; 37pp; English.  
 XX  
 CC The patent discloses antibodies or their fragments comprising a heavy  
 CC chain variable domain (VHH) derived from an immunoglobulin naturally  
 CC devoid of light chains specific for inhibiting human dietary enzymes. The  
 CC antibodies of the invention are useful for the preparation of medicaments  
 CC or food for inhibiting the activity of one or more human dietary enzymes  
 CC especially human pancreatic lipase (HPL) or human gastric lipase (HGL)  
 CC which are useful for the cosmetic control of body weight of human beings.  
 CC The present peptide sequence is HGL inhibiting VHH fragment, HGL #15 from  
 CC llama (camelid) species  
 XX  
 SQ Sequence 124 AA;  
 XX  
 Query Match 70.1%; Score 487; DB 4; Length 124;  
 Best Local Similarity 70.8%; Pred. No. 1.9e-32;  
 Matches 92; Conservative 12; Mismatches 20; Indels 6; Gaps 1;  
 QY 1 QVQLDSGGGLVQTGSLRLSCAASGPIGDVYLMGWYRQAPGKOREMVASITATGPPNYT 60  
 DB 1 QVQLDSGGGLVQAGSLRLSCAASGSISGMVYRQAPGKERELVARMSSDGTTSYT 60  
 QY 61 DSVKGRFTISRDNKNTYELQNNLKPEDTAVYYCNQAVRVRFSSDYTNKYGQGTQVTVS 120  
 DB 61 DSVKGRFTISRDNKNTYELQNNLKPEDTAVYYCNQAVRVRFSSDYTNKYGQGTQVTVS 114  
 QY 121 SEPKTPKQP 130  
 DB 115 SEPKTPKQP 124

61 DSVGRFTISRDNARGTVYIQMNSLKPEDTAVYVCAARSLELTPTSYDYWGOGTQTVTS 121

121 SEPTKRPQP 130  
|||||

121 SEPTKRPQP 130

RESULT 12  
ABG30620  
ID ABG30620 standard; protein; 131 AA.  
XX  
AC  
ABG30620;  
21-OCT-2002 (first entry)  
XX  
Immunoglobulin G specific heavy chain variable domain antibody #3.  
DE  
XX  
XX  
Heavy chain variable domain; antibody; protein array; aging; VHH;  
KW immunoglobulin; sticky phage-antibody; micro-panning; immune library;  
KW proteomic; mouse.  
OS  
Mus sp.  
XX  
WO200248193-A2.  
XX  
20-JUN-2002.  
XX  
03-DEC-2001; 2001WO-EP014471.  
XX  
13-DEC-2000; 2000EP-00311142.  
XX  
(UNIL ) UNILEVER NV.  
PA (UNIL ) UNILEVER PLC.  
PA (UNIL ) HINDUSTAN LEVER LTD.  
PI  
De Haard JW, Hermans P, Landa I, Verrips CT;  
XX  
WPI; 2002-593487/62.  
XX  
Novel protein array useful for detecting the presence of individual  
proteins in sample, comprises heavy-chain variable domain antibodies or  
antibody fragments obtainable from Camelidae.  
XX  
Disclosure; Fig 3; 80pp; English.  
XX  
The invention describes a protein array (I) comprising a number of heavy-  
chain variable domain antibodies or antibody fragments, obtainable from an  
Camelidae. The method is useful for removing abundant proteins from an  
extract or sample which do not provide useful information on the  
condition of a cell or tissue in the extract or sample to be  
investigated. (I) is useful for detecting the presence of individual  
proteins in a sample, comparing the distribution of proteins in different  
cell types, and identification of proteins that may be of importance in  
determining the altered properties of cells in disease, aging or other  
conditions. Using a heavy-chain variable domain derived from an  
immunoglobulin that is naturally devoid of light chains (VHH) in (I)  
provides a number of advantages, such as an improvement of  
sensitivity/resolution in the order of 10-100 times, and detection of  
post-translationally modified proteins. The invention also describes a  
method (II) that enables the simultaneous processing of large numbers of  
target antigens in a controlled way. The incorporated phage-ELISA  
generates on-line information about the success or failure of a certain  
panning condition. This feature combined with the microtiter plate format  
allows the complete automation of the technology, based on computer-made  
decisions on the values of the phage-ELISA for continuation of a limited  
number of selections. In (II), due to the fact that many different  
conditions can be tested, varying amounts of input-phages can be used  
simultaneously in order to decrease the enrichment of sticky phage-  
antibodies. Micro-panning is an effective tool for selecting both naive,  
synthetic and immune libraries on large numbers of different target  
molecules, enabling the generation of large panels of antibodies in short  
time frames needed for the generation of arrays (proteomics). This  
sequence represents an immunoglobulin (IgG) heavy chain variable domain

CC antibody (VHH)  
 SQ Sequence 131 AA;  
 Query Match. 68.4%; Score 475.5; DB 5; Length 131;  
 Best Local Similarity 73.3%; Pred. No. 1.7e-31;  
 Matches 96; Conservative 7; Mismatches 27; Indels 1; Gaps 1;

QY 1 QVQLDSGGGLVQGTGSLRLSCAASGPIGDVYLMGWYRQAPGKQREVMVASITATGPPNYT 60  
 DB 1 QVQLDSGGGLVQGTGSLRLSCAASGPIGDVYLMGWYRQAPGKQREVMVASITATGSGNYA 60  
 QY 61 DSVKGRFTISRDNKNTPEYLOMNLKPEDTAVYYCNAQ-VKRVFSSDYTYNWGGQTQVTV 119  
 DB 61 DSVKGRFTISRDNKNTPEYLOMNLKPEDTAVYYCNAETVRATTGRTITDLWGQTQVTV 120  
 QY 120 SSEPKTPKQP 130  
 DB 121 SSEPKTPKQP 131

RESULT 13  
 AAE05283  
 ID AAE05283 standard; protein; 152 AA.  
 XX  
 AC AAE05283;  
 XX  
 DT 18-SEP-2001 (first entry)  
 XX  
 DE Anti-potato SBEII (Clone46) VH region attached with myc and his6 tag.  
 XX  
 KW Potato; heavy chain immunoglobulin; pathogen resistance;  
 KW metabolism modulator; passive immunisation; heavy chain variable domain;  
 KW VH; anti-potato SBEII; starch branching enzyme; SBE A.  
 XX  
 OS Solanum tuberosum.  
 OS Unidentified.  
 OS Chimeric.  
 XX  
 PN EP1118669-A2.  
 XX  
 PD 25-JUL-2001.  
 XX  
 PF 08-DEC-2000; 2000EP-00310997.  
 XX  
 PR 17-DEC-1999; 99EP-00310188.  
 XX  
 PA (UNIL ) UNILEVER PLC.  
 PA (UNIL ) UNILEVER NV.  
 XX  
 PI Franken LGV, Van Der Logt CPE, Jobling SA, Teh Y;  
 XX  
 DR WPI; 2001-427157/46.  
 DR N-PSDB; AAD10054.  
 XX  
 PT Modifying a plant to produce an antibody useful for increasing pathogen  
 PT resistance or to modulate metabolism comprises introducing a DNA sequence  
 PT encoding a heavy chain immunoglobulin linked to a peptide that targets a  
 PT cellular compartment.  
 XX  
 PS Example 1; Fig 12; 81pp; English.  
 XX  
 CC The present invention relates to a method for modifying a plant to  
 CC produce an antibody or an active fragment or derivative, or a protein  
 CC functional equivalent, in a cellular compartment comprises introducing a  
 CC DNA sequence encoding a heavy chain immunoglobulin where the DNA is  
 CC linked to promoters and provided with an additional sequence encoding a  
 CC peptide capable of targeting heavy chain immunoglobulin to a cellular  
 CC compartment. The method is used for producing a heavy chain  
 CC immunoglobulin or an active fragment or derivative, or a protein that is  
 CC functionally equivalent for increasing the pathogen resistance in a plant  
 CC or to modulate metabolism in a plant. Under some circumstances it may be  
 CC desirable to retain the antibody product with the plant rather than

CC extracting and isolating the product. In particular, edible selected  
 CC antigens may be used in a method of passively immunising an animal,  
 CC preferably human, against the antigen, e.g., pathogenic organisms. The  
 CC present sequence is an anti-potato SBEII antibody (denoted Clone46) heavy  
 CC chain variable domain (VH) attached to peptide linkers, myc and his6 tag.  
 CC The potato SBEII is a starch branching enzyme also designated SBE A  
 SQ Sequence 152 AA;  
 Query Match. 67.9%; Score 472; DB 4; Length 152;  
 Best Local Similarity 72.5%; Pred. No. 3.9e-31;  
 Matches 95; Conservative 8; Mismatches 22; Indels 6; Gaps 2;

QY 1 QVQLDSGGGLVQGTGSLRLSCAASGPIGDVYLMGWYRQAPGKQREVMVASITATGPPNYT 60  
 DB 1 QVQLDSGGGLVQGTGSLRLSCAASGNTFSITAMAWYRQAPGKQREVMVASINSIGSTNYA 60  
 QY 61 DSVKGRFTISRDNKNTPEYLOMNLKPEDTAVYYCNA-QVVRFSSDYTYNWGGQTQVTV 119  
 DB 61 DSVKGRFTISRDNKNTPEYLOMNLKPEDTAVYYCAAGNLLVK-----RPYWGQTLTV 115  
 QY 120 SSEPKTPKQP 130  
 DB 116 SSEPKTPKQP 126

RESULT 14  
 AAE10561  
 ID AAE10561 standard; peptide; 129 AA.  
 XX  
 AC AAE10561;  
 XX  
 DT 10-DEC-2001 (first entry)  
 XX  
 DE HGL inhibiting VHH fragment, HGL #4 from llama species.  
 XX  
 KW Llama antibody; camelid; anorectic; heavy chain variable domain; VHH;  
 KW human dietary enzyme inhibitor; medicament; human pancreatic lipase; HPL;  
 KW food; human gastric lipase; HGL; cosmetic control; body weight.  
 XX  
 OS Lama sp.  
 XX  
 FH Key  
 FH Region 31..35  
 FT /label= CDR1  
 FT /note= "Complementarity determining region 1"  
 FT Region 50..64  
 FT /label= CDR2  
 FT /note= "Complementarity determining region 2"  
 FT Region 98..109  
 FT /label= CDR3  
 FT /note= "Complementarity determining region 3"  
 XX  
 PN EP1134231-A1.  
 XX  
 PD 19-SEP-2001.  
 XX  
 PF 20-FEB-2001; 2001EP-00200703.  
 XX  
 PR 14-MAR-2000; 2000EP-00200930.  
 XX  
 PA (UNIL ) UNILEVER NV.  
 PA (UNIL ) UNILEVER PLC.  
 XX  
 PI Bezemer S, Van De Burg M, De Haard JJW, Tarsilus E;  
 XX  
 DR WPI; 2001-572718/65.  
 XX  
 PT New antibody or its fragments for inhibiting human dietary enzymes,  
 PT useful for cosmetic control of body weight of human beings, comprises  
 PT heavy chain variable domain derived from immunoglobulin naturally devoid  
 PT of light chains.  
 XX

PS Example 4; Page 13; 37pp; English.

XX The patent discloses antibodies or their fragments comprising a heavy  
CC chain variable domain (VHH) derived from an immunoglobulin naturally  
CC devoid of light chains specific for inhibiting human dietary enzymes. The  
CC antibodies of the invention are useful for the preparation of medicaments  
CC or food for inhibiting the activity of one or more human dietary enzymes  
CC especially human pancreatic lipase (HPL) or human gastric lipase (HGL)  
CC which are useful for the cosmetic control of body weight of human beings.  
CC The present peptide sequence is HGL inhibiting VHH fragment, HGL #4 from  
CC llama (camelid) species

XX SQ Sequence 129 AA;

Query Match 67.3%; Score 467.5; DB 4; Length 129;  
Best Local Similarity 70.8%; Pred. No. 7.7e-31;  
Matches 92; Conservative 12; Mismatches 25; Indels 1; Gaps 1;

QY 1 QVQLDSGGGLVQTGGSLRLSCAASGPIGDVYLMGWYRQAPGKQREMVASITATGPPNYT 60

Db 1 QVQLDSGGGLVQAGGSLRLSCAASGTFNFAMGWYRQVPGKQRELVAIIGNDGATYYV 60

QY 61 DSVKGRFTISRNDKNTLEYLQNNLKPEDTAVYYCNAQVRVRFSSDYTNWGGQTQVTVS 120

Db 61 DSVKGRFTIARENAKNTVLYQSSLPEDTAVYYCKRGGLTQYSEH-DYWGQGTQVTVS 119

QY 121 SEPTKPKQP 130

Db 120 SEPTKPKQP 129

RESULT 15

AAE05288

ID AAE05288 standard; protein; 149 AA.

XX AC AAE05288;

DT 18-SEP-2001 (first entry)

DE Anti-potato SBEII (Clone68) VH region attached with myc and his6 tag.

XX Potato; heavy chain immunoglobulin; pathogen resistance;  
KW metabolism modulator; passive immunisation; heavy chain variable domain;  
KW VH; anti-potato SBEII; starch branching enzyme; SBE A.

XX Solanum tuberosum.

OS Unidentified.

OS Chimeric.

XX PF1118669-A2.

XX 25-JUL-2001.

XX 08-DEC-2000; 2000EP-00310997.

XX 17-DEC-1999; 99EP-00310188.

XX (UNIL ) UNILEVER PLC.

XX (UNIL ) UNILEVER NV.

XX Frenken IGJ, Van Der Logt CPE, Jobling SA, Teh Y;

XX WPI; 2001-427157/46.

XX N-PSDB; AAD10059.

XX Modifying a plant to produce an antibody useful for increasing pathogen  
PT resistance or to modulate metabolism comprises introducing a DNA sequence  
PT encoding a heavy chain immunoglobulin linked to a peptide that targets a  
PT cellular compartment.

XX Example 13; Fig 28; 81pp; English.

XX The present invention relates to a method for modifying a plant to

CC produce an antibody or an active fragment or derivative, or a protein  
CC functional equivalent, in a cellular compartment comprising introducing a  
CC DNA sequence encoding a heavy chain immunoglobulin, where the DNA is  
CC linked to promoters and provided with an additional sequence encoding a  
CC peptide capable of targeting heavy chain immunoglobulin to a cellular  
CC compartment. The method is used for producing a heavy chain  
CC immunoglobulin or an active fragment or derivative, or a protein that is  
CC functionally equivalent for increasing the pathogen resistance in a plant  
CC or to modulate metabolism in a plant. Under some circumstances it may be  
CC desirable to retain the antibody product with the plant rather than  
CC extracting and isolating the product. In particular, edible selected  
CC antigens may be used in a method of passively immunising an animal,  
CC preferably human, against the antigen, e.g., pathogenic organisms. The  
CC present sequence is an anti-potato SBEII antibody (denoted Clone68) heavy  
CC chain variable domain (VH) attached to peptide linkers, myc and his6 tag.  
CC The potato SBEII is a starch branching enzyme also designated SBE A

XX SQ Sequence 149 AA;

Query Match 67.0%; Score 465.5; DB 4; Length 149;

Best Local Similarity 72.3%; Pred. No. 1.3e-30;

Matches 94; Conservative 9; Mismatches 18; Indels 9; Gaps 1;

QY 1 QVQLDSGGGLVQTGGSLRLSCAASGPIGDVYLMGWYRQAPGKQREMVASITATGPPNYT 60

Db 3 QVQLDSGGGLVQAGGSLRLSCAASGTFNFAMGWYRQVPGKQRELVAIISAGRTWYA 62

QY 61 DSVKGRFTISRNDKNTLEYLQNNLKPEDTAVYYCNAQVRVRFSSDYTNWGGQTQVTVS 120

Db 63 DSVKGRFTISRDNKNTLEYLQNNLKPEDTAVYYCTAG-----GSYWGQGTQVTV 113

QY 121 SEPTKPKQP 130

Db 114 SEPTKPKQP 123

Search completed: October 5, 2004, 08:00:13

Job time : 99.9043 secs

This Page Blank (uspic)



Result No.	Score	Query		Length	DB	ID	Description
		Match	Length				
1	370.5	55.0	119	2	S36005		IG heavy chain V r
2	368.5	54.7	120	2	S55536		IG heavy chain V r
3	368	54.6	116	2	S31110		IG heavy chain - h
4	367	54.5	120	2	S48798		IG heavy chain V r
5	366.5	54.4	119	2	D38005		IG heavy chain V r
6	366.5	54.1	120	2	S55539		IG heavy chain V r
7	364.5	53.9	119	2	S31107		IG heavy chain - h
8	362.5	53.8	119	2	S31108		IG heavy chain - h
9	362	53.7	121	2	S55540		IG heavy chain V r
10	361.5	53.6	123	2	S32794		IG heavy chain V r
11	361.5	53.6	123	2	S31114		IG heavy chain - h
12	361.5	53.6	125	2	S30531		IG heavy chain V r
13	361.5	53.6	140	2	S31686		IG heavy chain V r
14	361.5	53.6	160	2	S05271		IG heavy chain V r
15	360.5	53.5	121	2	S28798		IG heavy chain V r
16	360	53.4	114	2	S31120		IG heavy chain - h
17	360	53.4	143	2	S23624		IG heavy chain V r
18	359.5	53.3	120	2	S55537		IG heavy chain V r
19	359.5	53.3	121	2	I55673		IG heavy chain - h
20	359	53.3	114	2	S46390		IG heavy chain V r
21	358.5	53.2	121	2	S31106		IG heavy chain - h
22	358.5	53.2	123	2	A36006		IG heavy chain V r
23	358.5	53.2	135	2	S31598		IG heavy chain V r
24	358	53.1	128	2	S28786		IG heavy chain V r
25	357.5	53.0	140	2	S31588		IG heavy chain V r
26	357	53.0	122	2	S31675		IG heavy chain V r
27	356.5	52.9	117	2	S78486		IG heavy chain V r
28	355.5	52.7	120	2	S55538		IG heavy chain V r
29	355.5	52.7	121	2	S19666		IG heavy chain V r

C:Superfamily: immunoglobulin V region; immunoglobulin homology  
C:Keywords: heterotetramer; immunoglobulin  
F:14-97/Domain: immunoglobulin homology <IMM>

Query Match 54.7%; Score 368.5; DB 2; Length 120;  
Best Local Similarity 65.0%; Pred. No. 3e-25;  
Matches 78; Conservative 11; Mismatches 30; Indels 1; Gaps 1;

QY 2 VQLQESGGGLVQAGGSLRLSCAASGSRISITMTWYRQAPGKERELVARMSSDGT-TSYT 60  
DB 1 VQLQESGGGLVQAGGSLRLSCAASGSRISITMTWYRQAPGKERELVARMSSDGT-TSYT 60  
QY 61 DSMKGRFTISRDNKNTVYLMNNLKPEDTAVYCKALISSYDGSWMDYGGQGTQVTVSS 120  
DB 61 DSVKGRFTISRDNKNTVYLMNNLKPEDTAVYCKALISSYDGSWMDYGGQGTQVTVSS 120

RESULT 3  
S31110  
Ig heavy chain - human  
C:Species: Homo sapiens (man)  
C:Date: 02-Dec-1993 #sequence\_revision 26-May-1995 #text\_change 17-Mar-1999  
C:Accession: S31110  
R:Raghuvaran, F.M.; Timmers, B.; Kenter, M.J.H.; van Tol, M.J.D.; Vossen, J.M.; Schuurman  
Eur. J. Immunol. 22, 247-251, 1992  
A:Title: Restricted utilization of germ-line V(H)3 genes and short diverse third complement  
A:Reference number: S31104; MUID:92111633; PMID:1730252  
A:Accession: S31110  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: mRNA  
A:Residues: 1-116 <RAA>  
A:Cross-references: EMBL:X62958  
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1991  
C:Superfamily: immunoglobulin V region; immunoglobulin homology  
C:Keywords: heterotetramer; immunoglobulin  
F:15-97/Domain: immunoglobulin homology <IMM>

Query Match 54.6%; Score 368; DB 2; Length 116;  
Best Local Similarity 63.3%; Pred. No. 3.2e-25;  
Matches 76; Conservative 14; Mismatches 26; Indels 4; Gaps 1;

QY 1 QVQLQESGGGLVQAGGSLRLSCAASGSRISITMTWYRQAPGKERELVARMSSDGTTSY 60  
DB 1 EVQLVETGGGLIQPGGSLRLSCAASGFTVSNYSYRQAPGKLEWVSYISGGSTYYA 60  
QY 61 DSMKGRFTISRDNKNTVYLMNNLKPEDTAVYCKALISSYDGSWMDYGGQGTQVTVSS 120  
DB 61 DSVKGRFTISRDNKNTVYLMNNLKPEDTAVYCKALISSYDGSWMDYGGQGTQVTVSS 116

RESULT 4  
S48798  
Ig heavy chain V region (anti-Sm, VH3/Dxp4/JH4b) - human  
C:Species: Homo sapiens (man)  
C:Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 23-Jul-1999  
C:Accession: S48798  
R:Mahmoudi, M.; Edwards, J.; Cairns, E.; Bell, D.  
submitted to the EMBL Data Library, October 1994  
A:Description: Molecular characterization of natural human anti-Sm autoantibodies.  
A:Reference number: S48797  
A:Accession: S48798  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-120 <MAH>  
A:Cross-references: EMBL:Z46382; NID:562324; PIDN:CAA86521.1; PID:G1340167  
C:Superfamily: immunoglobulin V region; immunoglobulin homology  
C:Keywords: heterotetramer; immunoglobulin  
F:15-98/Domain: immunoglobulin homology <IMM>

Query Match 54.5%; Score 367; DB 2; Length 120;  
Best Local Similarity 65.3%; Pred. No. 4.1e-25;  
Matches 79; Conservative 12; Mismatches 28; Indels 2; Gaps 2;

QY 1 QVQLQESGGGLVQAGGSLRLSCAASGSRISITMTWYRQAPGKERELVARMSSDGTTSY 59  
DB 1 EVQLQESGGGLVQAGGSLRLSCAASGSRISITMTWYRQAPGKERELVARMSSDGTTSY 60  
QY 60 TDSMKGRFTISRDNKNTVYLMNNLKPEDTAVYCKALISSYDGSWMDYGGQGTQVTVSS 119  
DB 61 ADSVKGRTISRDNKNTVYLMNNLKPEDTAVYCKALISSYDGSWMDYGGQGTQVTVSS 119

RESULT 5  
D36005  
Ig heavy chain V region (M43) - human  
C:Species: Homo sapiens (man)  
C:Date: 21-Dec-1990 #sequence\_revision 21-Dec-1990 #text\_change 16-Dec-1998  
C:Accession: D36005  
R:Schroeder Jr., H.W.; Wang, J.Y.  
Proc. Natl. Acad. Sci. U.S.A. 87, 6146-6150, 1990  
A:Title: Preferential utilization of conserved immunoglobulin heavy chain variable gene  
A:Reference number: A36005; MUID:90349571; PMID:2117273  
A:Accession: D36005  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-119 <SCH>  
A:Cross-references: GB:M34024  
A:Genetics: IGHD1  
A:Gene: GDB:IGH0; IGHDY1  
A:Cross-references: GDB:118731; OMIM:146910  
A:Map position: 14q32.33-14q32.33  
C:Superfamily: immunoglobulin V region; immunoglobulin homology  
C:Keywords: heterotetramer; immunoglobulin  
F:15-98/Domain: immunoglobulin homology <IMM>

Query Match 54.4%; Score 366.5; DB 2; Length 119;  
Best Local Similarity 64.5%; Pred. No. 4.4e-25;  
Matches 78; Conservative 13; Mismatches 27; Indels 3; Gaps 2;

QY 1 QVQLQESGGGLVQAGGSLRLSCAASGSRISITMTWYRQAPGKERELVARMSSDGTTSY 59  
DB 1 EVQLQESGGGLVQAGGSLRLSCAASGSRISITMTWYRQAPGKERELVARMSSDGTTSY 60  
QY 60 TDSMKGRFTISRDNKNTVYLMNNLKPEDTAVYCKALISSYDGSWMDYGGQGTQVTVSS 119  
DB 61 ADSVKGRTISRDNKNTVYLMNNLKPEDTAVYCKALISSYDGSWMDYGGQGTQVTVSS 118

QY 120 S 120  
DB 119 S 119

RESULT 6  
S55539  
Ig heavy chain V region pe24 - mouse (fragment)  
C:Species: Mus musculus (house mouse)  
C:Date: 27-Oct-1995 #sequence\_revision 03-Nov-1995 #text\_change 23-Jul-1999  
C:Accession: S55539  
R:Boettger, V.; Boettger, A.; Lane, E.B.; Spruce, B.A.  
J. Mol. Biol. 247, 932-946, 1995  
A:Title: Comprehensive epitope analysis of monoclonal anti-proenkephalin antibodies using  
A:Reference number: S55528; MUID:95239763; PMID:7536850  
A:Accession: S55539  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-120 <BOE>  
A:Cross-references: EMBL:X82593; NID:9854312; PIDN:CAA57929.1; PID:9854313  
C:Superfamily: immunoglobulin V region; immunoglobulin homology  
C:Keywords: heterotetramer; immunoglobulin  
F:14-97/Domain: immunoglobulin homology <IMM>

Query Match 54.1%; Score 364.5; DB 2; Length 120;  
Best Local Similarity 64.2%; Pred. No. 6.7e-25;  
Matches 77; Conservative 12; Mismatches 30; Indels 1; Gaps 1;

QY 2 VQLQESGGGLVQAGGSLRLSCAASGIRSIIMTWYRQAPGKERELVARMSDGT-TSYT 60  
DB 1 VQLQESGGGLVQAGGSLRLSCAASGIRSIIMTWYRQAPGKERELVARMSDGT-TSYT 60

QY 61 DSMKGRFTISRDNKNTVYLMNKLKPEDTAVYVC-KALISSYDGSWNDYGGQGTQTV 120  
DB 61 DSVKGRFTISRDNKNTVYLMNKLKPEDTAVYVC-KALISSYDGSWNDYGGQGTQTV 120

RESULT 7  
S31107  
IG heavy chain - human  
C/Species: Homo sapiens (man)  
C/Date: 02-Dec-1993 #sequence\_revision 26-May-1995 #text\_change 17-Mar-1999  
C/Accession: S31107  
R/Raaphorst, F.M.; Timmers, E.; Kenter, M.J.H.; van Tol, M.J.D.; Vossen, J.M.; Schuurman  
Eur. J. Immunol. 22, 247-251, 1992  
A/Title: Restricted utilization of germ-line V(H)3 genes and short diverse third complement  
A/Reference number: S31104; MUID:92111633; PMID:1730252  
A/Accession: S31107  
A/Status: preliminary; nucleic acid sequence not shown; translation not shown  
A/Molecule type: mRNA  
A/Residues: 1-119 <RAA>  
A/Cross-references: EMBL:X62955  
A/Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1991  
C/Superfamily: immunoglobulin V region; immunoglobulin homology  
C/Keywords: heterotetramer; immunoglobulin  
F/15-98/Domain: immunoglobulin homology <IMV>

Query Match 53.9%; Score 363.5; DB 2; Length 119;  
Best Local Similarity 66.4%; Pred. No. 8.1e-25;  
Matches 81; Conservative 12; Mismatches 24; Indels 5; Gaps 3;

QY 1 VQLQESGGGLVQAGGSLRLSCAASGIRSIIMTWYRQAPGKERELVARMS-SDGTTSY 59  
DB 1 EVQLVESGGGLVQPGGSLRLSCAASGFTFSYAMSWRQAPGKLEWVAISGSGSTYY 60

QY 60 TSMKGRFTISRDNKNTVYLMNKLKPEDTAVYVC-KALISSYDGSWNDYGGQGTQTV 118  
DB 61 ADSVKGRFTISRDNKNTVYLMNKLKPEDTAVYVC-KALISSYDGSWNDYGGQGTQTV 117

QY 119 SS 120  
DB 118 SS 119

RESULT 8  
S31108  
IG heavy chain - human  
C/Species: Homo sapiens (man)  
C/Date: 02-Dec-1993 #sequence\_revision 26-May-1995 #text\_change 17-Mar-1999  
C/Accession: S31108  
R/Raaphorst, F.M.; Timmers, E.; Kenter, M.J.H.; van Tol, M.J.D.; Vossen, J.M.; Schuurman  
Eur. J. Immunol. 22, 247-251, 1992  
A/Title: Restricted utilization of germ-line V(H)3 genes and short diverse third complement  
A/Reference number: S31104; MUID:92111633; PMID:1730252  
A/Accession: S31108  
A/Status: preliminary; nucleic acid sequence not shown; translation not shown  
A/Molecule type: mRNA  
A/Residues: 1-119 <RAA>  
A/Cross-references: EMBL:X62956  
A/Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1991  
C/Superfamily: immunoglobulin V region; immunoglobulin homology  
C/Keywords: heterotetramer; immunoglobulin  
F/15-98/Domain: immunoglobulin homology <IMV>

Query Match 53.8%; Score 362.5; DB 2; Length 119;  
Best Local Similarity 63.7%; Pred. No. 9.9e-25;  
Matches 79; Conservative 13; Mismatches 23; Indels 9; Gaps 3;

QY 1 QVQLQESGGGLVQAGGSLRLSCAASGIRSIIMTWYRQAPGKERELVARMS-SDGTTSY 59  
DB 1 EVQLVESGGGLVQPGGSLRLSCAASGFTFSYAMSWRQAPGKLEWVAISGSGSTYY 60

QY 60 TSMKGRFTISRDNKNTVYLMNKLKPEDTAVYVC-KALISSYDGSWNDYGGQGTQTV 116  
DB 61 ADSVKGRFTISRDNKNTVYLMNKLKPEDTAVYVC-KALISSYDGSWNDYGGQGTQTV 115

QY 117 TVSS 120  
DB 116 TVSS 119

RESULT 9  
S35540  
IG heavy chain V region pel3 - mouse (fragment)  
C/Species: Mus musculus (house mouse)  
C/Date: 27-Oct-1995 #sequence\_revision 10-Nov-1995 #text\_change 23-Jul-1999  
C/Accession: S55540  
R/Boettger, V.; Boettger, A.; Lane, E.B.; Spruce, B.A.  
J. Mol. Biol. 247, 932-946, 1995  
A/Title: Comprehensive epitope analysis of monoclonal anti-proenkephalin antibodies using  
utations in the variable region genes.  
A/Reference number: S55528; MUID:95239763; PMID:7536850  
A/Accession: S55540  
A/Status: preliminary  
A/Molecule type: mRNA  
A/Residues: 1-121 <BOE>  
A/Cross-references: EMBL:X82582; MID:g854288; PIDN:CAA57918.1; PID:g854289  
C/Superfamily: immunoglobulin V region; immunoglobulin homology  
C/Keywords: heterotetramer; immunoglobulin  
F/14-97/Domain: immunoglobulin homology <IMV>

Query Match 53.7%; Score 362; DB 2; Length 121;  
Best Local Similarity 63.6%; Pred. No. 1.1e-24;  
Matches 77; Conservative 14; Mismatches 28; Indels 2; Gaps 2;

QY 2 VQLQESGGGLVQAGGSLRLSCAASGIRSIIMTWYRQAPGKERELVARMSDGT-TSYT 60  
DB 1 VQLQESGGGLVQPGGSLRLSCAASGFTFSYAMCWVRQNFKEKLEWVATVSSSGTYYIS 60

QY 61 DSMKGRFTISRDNKNTVYLMNKLKPEDTAVYVC-KALISSYDGSW-NDYGGQGTQTV 119  
DB 61 DTVKGRFTISRDNKNTVYLMNKLKPEDTAVYVC-KALISSYDGSW-NDYGGQGTQTV 120

QY 120 S 120  
DB 121 S 121

RESULT 10  
S26794  
IG heavy chain V region - human  
C/Species: Homo sapiens (man)  
C/Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 17-Mar-1999  
C/Accession: S26794  
R/Mortari, F.; Newton, J.A.; Wang, J.Y.; Schroeder Jr., H.W.  
Eur. J. Immunol. 22, 241-245, 1992  
A/Title: The human cord blood antibody repertoire. Frequent usage of the V(H)7 gene faml  
A/Reference number: S26786; MUID:92111632; PMID:1730251  
A/Accession: S26794  
A/Status: preliminary  
A/Molecule type: mRNA  
A/Residues: 1-123 <MOR>  
A/Cross-references: EMBL:X61011  
C/Superfamily: immunoglobulin V region; immunoglobulin homology  
C/Keywords: heterotetramer; immunoglobulin  
F/15-98/Domain: immunoglobulin homology <IMV>

Query Match 53.6%; Score 361.5; DB 2; Length 123;  
Best Local Similarity 64.2%; Pred. No. 1.3e-24;  
Matches 79; Conservative 11; Mismatches 30; Indels 3; Gaps 2;

```

QY      1  QVQLQESGGGLVQAGGSLRLSCAASGSIISITMTWYRQAPGKERELVARMSDGTIT-SY 59
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      1  EVQLVESGGGLVQPGGSLRLSCAASGFTFSSYMNWVRQAPGKLEWVSIVSSSTIYY 60

QY      60  TDSMKGRFTISRDNKNTVYLHNKILKPDRTAVYYCALKLISYDGSW--NDYGGQGTQVT 117
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      61  ADSVKGRFTISRDNKNSLYLQNSLRDEDTAVYYCARSIKYIDENYYGMDVWGQGTITVT 120

QY      118  VSS 120
      |||
Db      121  VSS 123

RESULT 11
S31114
IG heavy chain - human
C:Species: Homo sapiens (man)
C:Date: 02-Dec-1993 #sequence_revision 26-May-1995 #text_change 17-Mar-1999
C:Accession: S31114
R:Raaphorst, F.M.; Timmers, E.; Kenter, M.J.H.; van Tol, M.J.D.; Vossen, J.M.; Schuurman
Eur. J. Immunol. 22, 247-251, 1992
A:Title: Restricted utilization of germ-line V(H)3 genes and short diverse third comple
A:Reference number: S31104; MUID:92111633; PMID:1730252
A:Accession: S31114
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: mRNA
A:Residues: 1-123 <PAA>
A:Cross-references: EMBL:X52963
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1991
C:Superfamily: immunoglobulin V region; immunoglobulin homology
C:Keywords: heterotrimer; immunoglobulin
F:15-98/Domain: immunoglobulin homology <IMM>

Query Match          53.6%; Score 361.5; DB 2; Length 123;
Best Local Similarity 65.9%; Pred. No. 1.3e-24;
Matches      81; Conservative 11; Mismatches 28; Indels 3; Gaps 3;

QY      1  QVQLQESGGGLVQAGGSLRLSCAASGSIISITMTWYRQAPGKERELVARMS-SDGTTSY 59
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      1  EVQLVESGGGLVQPGGSLRLSCAASGFTFSSYAMSVWRQAPGKLEWVSIVSSSGGSTYY 60

QY      60  TDSMKGRFTISRDNKNTVYLHNKILKPDRTAVYYC-KALKISYDGSW-NDYGGQGTQVT 117
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      61  ADSVKGRFTISRDNKNTLYLQNSLRDEDTAVYYCAKASLYLRFLEW.FDYWGQGTITVT 120

QY      118  VSS 120
      |||
Db      121  VSS 123

```

F;20-160/Product: Ig heavy chain (fragment) #status predicted <MAT>  
F;34-117/Domain: immunoglobulin homology <IMM>

```
Query Match          53.6%; Score 361.5; DB 2; Length 160;
Best Local Similarity 64.8%; Pred. No. 1.7e-24;
Matches 81; Conservative 12; Mismatches 27; Indels 5; Gaps 3;

QY 1 QVQLQESGGGLVQAGGSLRLSRAAGSIRSIIMTWYRQAPGKERELVARMS-SDGTTSY 59
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 20 EVQLVESGGGLVQPGGSLRLSRAAGSIRSIIMTWYRQAPGKERELVARMS-SDGTTSY 79
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 60 TDSMKGRFTISRDNKNTVYLHNNLKPEDTAVYVC-KALIS--SYDGSWNDYGGQGTQ 115
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 80 ADSVKGRFTISRDNKNTVYLHNNLKPEDTAVYVC-KALIS--SYDGSWNDYGGQGTQ 139
   |||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 116 VTVSS 120
   ||||
Db 140 VTVSS 144
   ||||
```

## RESULT 15

S26798  
Ig heavy chain V region - human  
C:Species: Homo sapiens (man)  
C:Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 20-Jun-2000  
C:Accession: S26798  
R:Mortari, F.; Newton, J.A.; Wang, J.Y.; Schroeder Jr., H.W.  
Eur. J. Immunol. 22, 241-245, 1992  
A:title: The human cord blood antibody repertoire. Frequent usage of the V(H)7 gene fami  
A:reference number: S26786; MUID:92111632; PMID:1730251  
A:Accession: S26798  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-121<MOR>  
A:Cross-references: EMBL:X61015; NID:g32795; PIDN:CAM43349.1; PID:g1335126  
C:Superfamily: immunoglobulin V region; immunoglobulin homology  
C:Keywords: heterotetramer; immunoglobulin  
F;15-98/Domain: immunoglobulin homology <IMM>

```
Query Match          53.5%; Score 360.5; DB 2; Length 121;
Best Local Similarity 63.6%; Pred. No. 1.5e-24;
Matches 77; Conservative 10; Mismatches 33; Indels 1; Gaps 1;

QY 1 QVQLQESGGGLVQAGGSLRLSRAAGSIRSIIMTWYRQAPGKERELVARMSDGT-TSY 59
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 EVQLVESGGGLVQPGGSLRLSCTASGFNFSHMHVVRQAPGKGLVWVSHIKYGGTYNY 60
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 60 TDSMKGRFTISRDNKNTVYLHNNLKPEDTAVYVC-KALISSYDGSWNDYGGQGTQTVTS 119
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 61 ADSVKGRFTISRDNKNTVYLHNNLKPEDTAVYVCRAIGRYSYGYLDYWGQGTLTVTS 120
   |||:|||||:|||||:|||||:|||||:|||||:

QY 120 S 120
   |
Db 121 S 121
   |
```

Search completed: October 5, 2004, 08:17:17  
Job time : 21.9028 secs

This Page Blank (uspto)

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2004, 07:29:35 ; Search time 11.5463 Seconds  
(without alignments)  
581.749 Million cell updates/sec

Title: US-09-805-290A-25

Perfect score: 674

Sequence: 1 QVQLQESGGLVQAGGSLRL.....GGQGTQTVTSSEPKTPKQP 129

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	334	49.6	126	1 HV3K HUMAN	P01772 homo sapien
2	331	49.1	116	1 HV3K HUMAN	P01781 homo sapien
3	330.5	49.0	115	1 HV3F HUMAN	P01767 homo sapien
4	326.5	48.4	119	1 HV3I HUMAN	P01770 homo sapien
5	326	48.4	116	1 HV05 CARAU	P19181 carassius a
6	325.5	48.3	115	1 HV3D HUMAN	P01765 homo sapien
7	325	48.2	119	1 HV40 MOUSE	P01810 mus musculus
8	324.5	48.1	117	1 HV3C HUMAN	P01764 homo sapien
9	320	47.5	122	1 HV3E HUMAN	P01768 homo sapien
10	320	47.5	136	1 HV1C MOUSE	P01783 mus musculus
11	319.5	47.4	117	1 HV02 CANFA	P01785 canis famul
12	319	47.3	119	1 HV37 MOUSE	P01807 mus musculus
13	318	47.2	114	1 HV3B HUMAN	P01763 homo sapien
14	317	47.0	119	1 HV38 MOUSE	P01808 mus musculus
15	315	46.7	120	1 HV3U HUMAN	P01782 homo sapien
16	313.5	46.5	119	1 HV3I HUMAN	P01773 homo sapien
17	313	46.4	122	1 HV3A HUMAN	P01762 homo sapien
18	312.5	46.4	120	1 HV3E HUMAN	P01766 homo sapien
19	312.5	46.4	142	1 HV01 RAT	P01805 rattus norv
20	311.5	46.2	115	1 HV33 MOUSE	P01802 mus musculus
21	310.5	46.1	118	1 HV39 MOUSE	P01809 mus musculus
22	309	45.8	114	1 HV01 CANFA	P01784 canis famul
23	308	45.7	122	1 HV3H HUMAN	P01769 homo sapien
24	307.5	45.6	123	1 HV2S MOUSE	P01794 mus musculus
25	305.5	45.3	115	1 HV32 MOUSE	P01801 mus musculus
26	305.5	45.3	117	1 HV53 MOUSE	P18524 mus musculus
27	305	45.3	119	1 HV3M HUMAN	P01774 homo sapien
28	304.5	45.2	121	1 HV3J HUMAN	P01771 homo sapien
29	304	45.1	97	1 HV56 MOUSE	P18527 mus musculus
30	303.5	45.0	113	1 HV30 MOUSE	P01799 mus musculus
31	303.5	45.0	117	1 HV55 MOUSE	P18526 mus musculus
32	303	45.0	119	1 HV3F HUMAN	P01777 homo sapien
33	302	44.8	122	1 HV2I MOUSE	P01790 mus musculus

34	301.5	44.7	123	1 HV22 MOUSE	P01791 mus musculus
35	301	44.7	122	1 HV20 MOUSE	P01789 mus musculus
36	299.5	44.4	123	1 HV24 MOUSE	P01793 mus musculus
37	299	44.4	111	1 HV35 MOUSE	P01804 mus musculus
38	298.5	44.3	113	1 HV27 MOUSE	P01796 mus musculus
39	298.5	44.3	123	1 HV18 MOUSE	P01787 mus musculus
40	298	44.2	117	1 HV30 HUMAN	P01776 homo sapien
41	298	44.2	119	1 HV3N HUMAN	P01775 homo sapien
42	297.5	44.1	117	1 HV54 MOUSE	P18525 mus musculus
43	296.5	44.0	98	1 HV57 MOUSE	P18528 mus musculus
44	295.5	43.8	117	1 HV41 MOUSE	P01811 mus musculus
45	295.5	43.8	123	1 HV23 MOUSE	P01792 mus musculus

#### ALIGNMENTS

```

RESULT 1
HV3K HUMAN
ID HV3K HUMAN STANDARD; PRT; 126 AA.
AC P01772
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE IG heavy chain V-III region KOL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_taxid=9606;
RN [1]
RP SEQUENCE, AND DISULFIDE BONDS.
RX MEDLINE=83289131; PubMed=6884994;
RA Schmidt W.E., Jung H.-D., Palm W., Hilschmann N.;
RT "Three-dimensional structure determination of antibodies. Primary
RT structure of crystallized monoclonal immunoglobulin IgG1 KOL, I.";
RL Hoppe-Seyler's Z. Physiol. Chem. 364:713-747(1983).
RN [2]
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).
RX MEDLINE=81072295; PubMed=7441755;
RA Marquart M., Deisenhofer J., Huber R., Palm W.;
RT "Crystallographic refinement and atomic models of the intact
RT immunoglobulin molecule KOL and its antigen-binding fragment at 3.0 A
RT and 1.0-A resolution.";
RL J. Mol. Biol. 141:369-391(1980).
CC -! SIMILARITY: Contains 1 immunoglobulin-like domain.
DR PIR; A02055; GIHUKL.
DR PDB; 2PB4; 12-JUL-89.
DR PDB; 2IG2; 12-JUL-89.
DR GO; GO:0005576; C:extracellular; NAS.
DR GO; GO:0003823; F:antigen binding; NAS.
DR GO; GO:0006955; P:immune response; NAS.
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003596; IG_v.
DR Pfam; PF00047; Ig; 1_v.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS00835; IG-LIKE; 1.
KW Immunoglobulin v region; 3D-structure; Pyrrolidone carboxylic acid.
FT DOMAIN 1 112
FT MOD RES 1 1
FT DISULFID 22 96
FT DISULFID 105 110
FT STRAND 3 7
FT STRAND 11 12
FT TURN 14 15
FT STRAND 18 25
FT HELIX 29 31
FT STRAND 34 39
FT TURN 41 42
FT STRAND 45 51
FT TURN 53 54
FT STRAND 58 60
FT HELIX 62 64
FT STRAND 65 65

```

FT TURN 66 67  
FT STRAND 68 73  
FT TURN 74 77  
FT STRAND 78 83  
FT STRAND 88 90  
FT HELIX 88 90  
FT STRAND 92 99  
FT STRAND 106 106  
FT TURN 107 108  
FT STRAND 109 109  
FT STRAND 113 116  
FT STRAND 120 124  
FT STRAND 126 126  
FT NON TER 126 126  
SQ SEQUENCE 126 AA; 13718 MW; E4D71B52B16F8776 CRC64;

Query Match 49.6%; Score 334; DB 1; Length 126;  
Best Local Similarity 58.7%; Pred. No. 8.2e-26;  
Matches 74; Conservative 13; Mismatches 33; Indels 6; Gaps 2;  
QY 1 QVQLVDSGGGLVQAGGSLRLSCAASGSRISITMTWYRQAPGKERELVARMSSDGTIS-Y 59  
DB 1 QVQLVESGGGVQPGESLRLSCAASGSRIFSSYMYWYRQAPGKLEWVLIWDDGSDQHY 60  
QY 60 TDSMKGRFTISRDNKNTVYLHNNLKPEDTAVYVC-----KALISSYDGSWYDGGQGT 114  
DB 61 ADSVKGRFTISRDNKNTVYLHNNLKPEDTAVYFCARDGCHGFCSSASCDFPDYWGQGT 120  
QY 115 QVTSS 120  
DB 121 PVTSS 126

RESULT 2  
HV3I HUMAN STANDARD; PRT; 116 AA.  
ID HV3I HUMAN  
AC P01781;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Ig heavy chain V-III region GAL.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE.  
RX MEDLINE=75059123; PubMed=4803843;  
RA Watanabe S., Barnikel H.U., Horn J., Bertram J., Hilschmann N.;  
RT "The primary structure of a monoclonal IgM-immunoglobulin  
(macroglobulin Gal.), II: the amino acid sequence of the H-chain (mu-  
type), subgroup H II. Architecture of the complete IgM-molecule.";  
RL Hoppe-Seyler's Z. Physiol. Chem. 354:1505-1509 (1973).  
RN [2]  
RP REVISION TO 28-33.  
RA Hilschmann N.;  
RL Submitted (JUN-1975) to the PIR data bank.  
CC -!- MISCELLANEOUS: THIS MU CHAIN WAS ISOLATED FROM A WALDENSTROM'S  
MACROGLOBULIN.  
CC -!- SIMILARITY: Contains 1 immunoglobulin-like domain.  
DR PIR; A02064; M3HUGL.  
DR HSSP; P01772; 2F3A.  
DR GO; GO:0005576; C:extracellular; NAS.  
DR GO; GO:0003823; F:antigen binding; NAS.  
DR GO; GO:0006955; P:immune response; NAS.  
DR InterPro; IPR007110; IG-like.  
DR InterPro; IPR003596; IG\_V.  
DR Pfam; PF00047; Ig; 1.  
DR SMART; SM00406; Igv; 1.  
DR PROSITE; PS50835; IG LIKE; 1.  
KW Immunoglobulin V region.  
FT DOMAIN 1 112 IG-LIKE.  
FT NON TER 116  
SQ SEQUENCE 116 AA; 12730 MW; 2C67CA9AAAAA1282 CRC64;

Query Match 49.1%; Score 331; DB 1; Length 116;  
Best Local Similarity 59.3%; Pred. No. 1.5e-25;  
Matches 73; Conservative 11; Mismatches 29; Indels 10; Gaps 3;  
QY 1 QVQLVESGGGLVQAGGSLRLSCAASGSRISITMTWYRQAPGKERELVARMSSDGT-TSY 59  
DB 1 EVQLVESGGDLVQPGESLRLSCAASGSRFBLGTMWYRQAPGKLEWVNIKZBSZBY 60  
QY 60 TDSMKGRFTISRDNKNTVYLHNNLKPEDTAVYCKALISSYDGSW--NDYGGQGTQVT 117  
DB 61 VDSVKGRFTISRDNKNTVYLHNNLKPEDTAVYCAR-----GWGGGYWGQGTIVT 113  
QY 118 VSS 120  
DB 114 VST 116

RESULT 3  
HV3I HUMAN STANDARD; PRT; 115 AA.  
ID HV3I HUMAN  
AC P01767;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Ig heavy chain V-III region BUT.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE.  
RX MEDLINE=78137069; PubMed=416441;  
RA Torano A., Putnam F.W.;  
RT "Complete amino acid sequence of the alpha 2 heavy chain of a human  
IgA2 immunoglobulin of the A2m (2) allotype.";  
RL Proc. Natl. Acad. Sci. U.S.A. 75:966-969 (1978).  
CC -!- MISCELLANEOUS: THE SEQUENCE OF THE ALPHA-2, A2M(2) ALLOTYPE, C  
REGION OF THIS MYELOMA PROTEIN IS ALSO GIVEN.  
CC -!- SIMILARITY: Contains 1 immunoglobulin-like domain.  
DR PIR; A02050; A2HUBU.  
DR HSSP; P01789; 1MCP.  
DR GO; GO:0005576; C:extracellular; NAS.  
DR GO; GO:0003823; F:antigen binding; NAS.  
DR GO; GO:0006955; P:immune response; NAS.  
DR InterPro; IPR007110; IG-like.  
DR InterPro; IPR003596; IG\_V.  
DR Pfam; PF00047; Ig; 1.  
DR SMART; SM00406; Igv; 1.  
DR PROSITE; PS50835; IG LIKE; 1.  
KW Immunoglobulin V region.  
FT DOMAIN 1 111 IG-LIKE.  
FT NON TER 115  
SQ SEQUENCE 115 AA; 12379 MW; 208876A7DF52DCF4 CRC64;

Query Match 49.0%; Score 330.5; DB 1; Length 115;  
Best Local Similarity 57.5%; Pred. No. 1.6e-25;  
Matches 69; Conservative 18; Mismatches 28; Indels 5; Gaps 1;  
QY 1 QVQLVESGGGLVQAGGSLRLSCAASGSRISITMTWYRQAPGKERELVARMSSDGTTSYT 60  
DB 1 EVQLVETGGGLVQPGESLRLSCAASGSRFVSBHNSWVRQAFKALZWVSATYRGTTYYA 60  
QY 61 DSMKGRFTISRDNKNTVYLHNNLKPEDTAVYCKALISSYDGSWYDGGQGTQVTSS 120  
DB 61 DSVKGRFTISRDNKNTVYLHNNLKPEDTAVYCAR-----ARDLAAARLFKGTIVTSS 115  
RESULT 4  
HV3I HUMAN STANDARD; PRT; 119 AA.  
ID HV3I HUMAN  
AC P01770;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)



```

DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ig heavy chain V-III region NIE.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE.
RA Ponsstingl H., Hilschmann N.;
RX MEDLINE=77070269; PubMed=826475;
RT "The rule of antibody structure. The primary structure of a
RT monoclonal IgG1 immunoglobulin (myeloma protein Nie). III. The
RT chymotryptic peptides of the H-chain, alignment of the tryptic
RT peptides and discussion of the complete structure."
RL Hoppe-Seyler's Z. Physiol. Chem. 357:1571-1604(1976).
RN [2]
RP DISULFIDE BOND
RX MEDLINE=77070267; PubMed=1002129;
RA Dreker L., Schwarz J., Reichel W., Hilschmann N.;
RT "Rule of antibody structure. The primary structure of a monoclonal
RT IgG1 immunoglobulin (myeloma protein Nie). I: Purification and
RT characterization of the protein, the L- and H-chains, the
RT cyanogen bromide cleavage products, and the disulfide bridges."
RL Hoppe-Seyler's Z. Physiol. Chem. 357:1515-1540(1976).
CC -!- MISCELLANEOUS: THIS CHAIN WAS ISOLATED FROM AN IGG1 MYELOMA
CC PROTEIN.
CC -!- SIMILARITY: Contains 1 immunoglobulin-like domain.
DR FIR; A91668; GIHUNI.
DR HSP; P01772; 2FB4.
DR GO; GO:0005576; C:extracellular; NAS.
DR GO; GO:0003823; F:antigen binding; NAS.
DR GO; GO:0009555; P:immune response; NAS.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003596; Ig_v.
DR SMART; SM00406; IGV; 1.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS50835; IG LIKE; 1.
KW Immunoglobulin V region; Pyridolone carboxylic acid.
FT DOMAIN 1 112 IG-LIKE.
FT MOD RES 1 1 PYRROLIDONE CARBOXYLIC ACID.
FT DISULFID 22 96
FT NON TER 119 119
SQ SEQUENCE 119 AA; 13242 MW; C96935A6E5E165B CRC64;

Query Match 48.4%; Score 326.5; DB 1; Length 119;
Best Local Similarity 58.5%; Pred. No. 4.1e-25;
Matches 72; Conservative 15; Mismatches 29; Indels 7; Gaps 3;

QY 1 QVQLQESGGGLVQAGGSLRLSCAASGSRISIMTWYRQAPGKERELVARMSSDGTTS-Y 59
Db 1 QVQLVCSGGGVQPGESLRLSCAASGTFPSRTIHWVQAFGKGLWAVVMSYGBBKHY 60

QY 60 TDSMKGRFTISRDNKNTVYLHNNLKPEDTAVYCKAL--ISSYDGSWNDYGGGQTQVT 117
Db 61 ADSVNGRFTISRDNKNTVYLNNNSLRPEDTAVYCARIRDTPAFHW---GGGTLVT 116

QY 118 VSS 120
Db 117 VSS 119

RESULT 5
ID HV05 CARAU STANDARD; PRT; 116 AA.
AC P19181.
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DE 15-JUL-1999 (Rel. 38, Last annotation update)
DE Ig heavy chain V region 5A precursor.
OS Carassius auratus (Goldfish).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Carassius.

```

```

OX NCBI_TaxID=7957;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88144476; PubMed=3125551;
RA Wilson M.R., Middleton D., Warr G.W.;
RT "Immunoglobulin heavy chain variable region gene evolution: structure
RT and family relationships of two genes and a pseudogene in a teleost
RT fish."
RL Proc. Natl. Acad. Sci. U.S.A. 85:1566-1570(1988).
DR HSP; P01772; 2FB4.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003596; Ig_v.
DR Pfam; PF00047; Ig; 1.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS50835; IG LIKE; 1.
KW Immunoglobulin V region; Signal.
FT SIGNAL 1 19
FT CHAIN 20 116 IG HEAVY CHAIN V REGION 5A.
FT DOMAIN 20 49 FRAMEWORK-1.
FT DOMAIN 50 54 COMPLEMENTARITY-DETERMINING-1.
FT DOMAIN 55 68 FRAMEWORK-2.
FT DOMAIN 69 84 COMPLEMENTARITY-DETERMINING-2.
FT DOMAIN 85 116 FRAMEWORK-3.
FT DISULFID 41 114 BY SIMILARITY.
FT NON TER 116 116
SQ SEQUENCE 116 AA; 12808 MW; 9C2279E2DF199B12 CRC64;

Query Match 48.4%; Score 326; DB 1; Length 116;
Best Local Similarity 67.4%; Pred. No. 4.5e-25;
Matches 64; Conservative 12; Mismatches 19; Indels 0; Gaps 0;

QY 1 QVQLQESGGGLVQAGGSLRLSCAASGSRISIMTWYRQAPGKERELVARMSSDGTTSY 60
Db 20 EVQLVESGGGLIQPGSLRLSCAASGTFVSSNTSMWVRQPGKGLWVSVIYSGSTYYA 79

QY 61 DSMKGRFTISRDNKNTVYLHNNLKPEDTAVYVC 95
Db 80 DSVAGRFTISRDNKNTVYLQMNLSRAEDTAVYVC 114

RESULT 6
ID HV3D HUMAN STANDARD; PRT; 115 AA.
AC P01765;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ig heavy chain V-III region TIL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE.
RX MEDLINE=78005528; PubMed=409716;
RA Wang A.-C., Wang J.Y., Fudenberg H.H.;
RT "Immunoglobulin structure and genetics. Identity between variable
RT regions of a mu and a gamma2 chain."
RL J. Biol. Chem. 252:7192-7199(1977).
CC -!- MISCELLANEOUS: THE SEQUENCES OF THE V REGIONS OF THE HEAVY CHAINS
CC OF IGM AND IGG2 ISOLATED FROM A SINGLE PATIENT WITH BICLONAL
CC GAMMOPATHY ARE IDENTICAL. THEIR LIGHT CHAINS ARE APPARENTLY ALSO
CC IDENTICAL.
CC -!- SIMILARITY: Contains 1 immunoglobulin-like domain.
DR FIR; A02048; H3HUTL.
DR HSP; P01772; 2FB4.
DR GO; GO:0005576; C:extracellular; NAS.
DR GO; GO:0003823; F:antigen binding; NAS.
DR GO; GO:0006955; P:immune response; NAS.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003596; Ig_v.
DR Pfam; PF00047; Ig; 1.
DR SMART; SM00406; IGV; 1.

```

```

DR PROSITE; PS0835; IG LIKE; 1.
KW Immunoglobulin V region.
FT DOMAIN 1 108 IG-LIKE.
FT NON TER 115
SQ SEQUENCE 115 AA; 12356 MW; 4DCC67D179F62326 CRC64;

Query Match
  48.3%; Score 325.5; DB 1; Length 115;
Best Local Similarity 58.2%; Pred. No. 5e-25;
Matches 71; Conservative 17; Mismatches 25; Indels 9; Gaps 4;

QY 1 QVQLDSGGGLVQAGGSLRLSCAASGIRSIISIMTWYRQAPGKERELVARMSS-DGTTSY S9
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 1 EVQLDSGGGLVQPGGSLKLSCAASGDFPSKYMSWVRQAPGKGLWVGIZGLSVSZSY 60
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 60 TDSMKGRFTISRDNKNTVYLHNNLKPEDTAVYCKALISSYDGSNDYGGGQTQVTV 118
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 61 ABSVGRFTISRDSKNT-----VNSLRADTAVYCAKGVSAV---YFYWGZGTLVTV 113
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 119 SS 120
   ||
DB 114 SS 115

RESULT 7
HV40 MOUSE STANDARD; PRT; 119 AA.
AC P01810;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Ig heavy chain V region J539.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP PRELIMINARY SEQUENCE.
RX MEDLINE=79223895; PubMed=111245;
RA Rao D.N., Rudikoff S., Krutzsch H., Potter M.;
RT "Structural evidence for independent joining region gene in
RT immunoglobulin heavy chains from anti-galactan myeloma proteins and
RT its potential role in generating diversity in
RT complementarity-determining regions.";
RL proc. Natl. Acad. Sci. U.S.A. 76:2890-2894 (1979).
RN [2]
RP X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS).
RX MEDLINE=88217852; PubMed=3449853;
RA Suh S.W., Bhat T.N., Navia M.A., Cohen G.H., Rao D.N., Rudikoff S.,
RA Davies D.R.;
RT "The galactan-binding immunoglobulin Fab J539: an X-ray diffraction
RT study at 2.74-A resolution.";
RL Proteins 1:74-80(1986).
CC -!- MISCELLANEOUS: THIS CHAIN WAS ISOLATED FROM A MYELOMA PROTEIN THAT
CC BINDS GALACTAN.
DR PIR; A02080; AVMSJ5.
DR PDB; 2FEJ; 15-OCT-90.
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003596; IG_V.
DR Pfam; PF00047; IG_1.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS0835; IG LIKE; 1.
KW Immunoglobulin V region; 3D-structure.
FT NON TER 119
FT STRAND 3 7
FT STRAND 10 12
FT TURN 14 15
FT STRAND 18 25
FT HELIX 29 31
FT STRAND 34 39
FT TURN 41 42
FT STRAND 45 51
FT TURN 53 54
FT TURN 55 56
FT STRAND 58 60

us-09-805-290a-25.rsp
Page 4

```

```
FT NON TER 117 117
SQ SEQUENCE 117 AA; 12582 MW; E8267333F1A3CB0F1 CRC64;

Query Match 48.1%; Score 324.5; DB 1; Length 117;
Best Local Similarity 69.8%; Pred. No. 6.3e-25;
Matches 67; Conservative 10; Mismatches 18; Indels 1; Gaps 1;

QY 1 QVQLQESGGGLVQAGGSLRLSCAASGSRISITMTWYRQAPGKERELVARMS-SDGTTSY 59
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 20 EVQLLEGGGGLVQPGGSLRLSCAASGFTTSSYAMSWRQAPGKLEWVAISGSGSTYY 79
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 60 TDSMKGRFTISRDNKNTVYLNHNKLPEDTAVYYC 95
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 80 GDSVKGRFTISRDNKNTVYLNHNKLPEDTAVYYC 115
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 9
HV3G_HUMAN STANDARD; PRT; 122 AA.
ID HV3G_HUMAN
AC P01768;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DE 10-OCT-2003 (Rel. 42, Last annotation update)
DE IG heavy chain V-III region CAM.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE
RX MEDLINE=81013859; PubMed=6774332;
RA Lehman D.W., Putnam F.W.;
RT "Amino acid sequence of the variable region of a human mu chain: location of a possible JH segment."
RL Proc. Natl. Acad. Sci. U.S.A. 77:3239-3243 (1980).
CC -I- MISCELLANEOUS: THIS MU CHAIN WAS ISOLATED FROM THE PLASMA OF A PATIENT WITH MACROGLOBULINEMIA.
CC -I- SIMILARITY: Contains 1 immunoglobulin-like domain.
DR PIR; A02051; M3HUAM.
DR HSSP; P01772; 2FB4.
DR GO; GO:0005576; C:extracellular; NAS.
DR GO; GO:0003823; P:antigen binding; NAS.
DR GO; GO:0006955; P:immune response; NAS.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003596; Ig_V.
DR SMART; SM00406; IG; 1.
DR PROSITE; PS00835; IG LIKE; 1.
FT DOMAIN 1 112 IG-LIKE.
FT MOD_RES 1 112 PYRROLIDONE CARBOXYLIC ACID.
FT NON_TER 122 122
SQ SEQUENCE 122 AA; 13668 MW; A42D0F17D252F1C2 CRC64;

Query Match 47.5%; Score 320; DB 1; Length 122;
Best Local Similarity 56.5%; Pred. No. 1.8e-24;
Matches 70; Conservative 16; Mismatches 32; Indels 6; Gaps 3;

QY 1 QVQLQESGGGLVQAGGSLRLSCAASGSRISITMTWYRQAPGKERELVARMSDGTTS-Y 59
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 QVELVESGGGVZPGRSLRLSCAASGFTFSNYAMHWVRQPGKLEWVAISYGBBKYY 60
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 60 TDSMKGRFTISRDNKNTVYLNHNKLPEDTAVYYC---KALISSYDGSNDYGGGQTV 116
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 61 ASVVKGRFTISRDNKNTVYLNHNKLPEDTAVYYC---KALISSYDGSNDYGGGQTV 118
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 117 TVSS 120
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 119 TVSS 122
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 10
HV16_MOUSE STANDARD; PRT; 136 AA.
ID HV16_MOUSE
AC P01783;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DE 10-OCT-2003 (Rel. 42, Last annotation update)
DE IG heavy chain V region MOPC 21 precursor (Fragment).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=81234548; PubMed=6788376;
RA Bothwell A.L.M., Paskind M., Reth M., Imanishi-Kari T., Rajewsky K., Baltimore D.;
RT "Heavy chain variable region contribution to the NPb family of antibodies: somatic mutation evident in a gamma 2a variable region.";
RL Cell 24:625-637 (1981).
RN [2]
RP SEQUENCE OF 17-136.
RX MEDLINE=77100368; PubMed=401950;
RA Aderugbo K., Milstein C., Secher D.S.;
RT "Molecular analysis of spontaneous somatic mutants."
RL Nature 265:299-304 (1977).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@sib-sib.ch).
CC -----
DR EMBL; J00522; AAD15290.1; -.
DR PIR; E30809; G1MS21.
DR PDB; 1IGC; 02-JUN-95.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003596; Ig_V.
DR Pfam; PF00047; Ig; 1.
DR SMART; SM00406; IG; 1.
DR PROSITE; PS00835; IG LIKE; 1.
KW Immunoglobulin V region; Signal; 3D-structure.
FT NON_TER 1 1
FT SIGNAL <1 16
FT CHAIN 17 136 IG HEAVY CHAIN V REGION MOPC 21.
FT DOMAIN 115 119 D SEGMENT.
FT DOMAIN 120 136 JH4 SEGMENT.
FT DISULFID 38 112
FT CONFLICT 75 78 HYAD -> DYAH (IN REF. 2).
FT CONFLICT 89 90 DN -> ND (IN REF. 2).
FT CONFLICT 115 115 W -> H (IN REF. 2).
FT CONFLICT 120 120 Y -> W (IN REF. 2).
FT NON_TER 136 136
SQ SEQUENCE 136 AA; 15071 MW; 2276A98DBDF7016 CRC64;

Query Match 47.5%; Score 320; DB 1; Length 136;
Best Local Similarity 60.8%; Pred. No. 2.1e-24;
Matches 73; Conservative 11; Mismatches 34; Indels 2; Gaps 2;

QY 2 VQLQESGGGLVQAGGSLRLSCAASGSRISITMTWYRQAPGKERELVARMSDGTG-TSY 60
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 18 VQLVESGGGLVQPGGSRKLSCAASGFTFSFGHWHVQAPKLEWVAISSGSSTLHYA 77
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 61 DSMKGRFTISRDNKNTVYLNHNKLPEDTAVYYCKALISSYDGSNDYGGGQTVTVSS 120
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 78 DTVKGRFTISRDNKNTVYLNHNKLPEDTAVYYC-ARWGNPYFYAMDYWGQSTVTVSS 136
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 11
HV02_CANFA STANDARD; PRT; 117 AA.
ID HV02_CANFA
AC P01785;
DT 21-JUL-1986 (Rel. 01, Created)
```

```

DR 21-JUL-1986 (Rel. 01, Last sequence update)
DR 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ig heavy chain V region MOO.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
CX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE OF 1-112.
RX MEDLINE=72242268; PubMed=407924;
RA Wasserman R.L.; Capra J.D.;
RT "Primary structure of the variable regions of two canine
RT immunoglobulin heavy chains."
RL Biochemistry 16:3160-3168(1977).
RN [2]
RP SEQUENCE OF 113-117.
RX MEDLINE=80077682; PubMed=117299;
RA McCumber L.J.; Capra J.D.;
RT "The complete amino-acid sequence of a canine mu chain."
RL Mol. Immunol. 16:565-570(1979).
CC -!- MISCELLANEOUS: THIS MU CHAIN WAS ISOLATED FROM A MYELOMA PROTEIN.
CC -!- SIMILARITY: Contains 1 immunoglobulin-like domain.
DR PIR; A90403; MHDGMO.
DR HSSP; P01772; 2FB4.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003596; Ig_v.
DR Pfam; PF00047; Ig; 1.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS50835; IG LIKE; 1.
KW Immunoglobulin V region.
FT DOMAIN 1 116
FT NON_TER 117
FT NCBI_TaxID=9615;
SQ SEQUENCE 117 AA; 12703 MW; FCE0730950A84B35 CRC64;

Query Match 47.4%; Score 319.5; DB 1; Length 117;
Best Local Similarity 58.3%; Pred. No. 2e-24;
Matches 70; Conservative 15; Mismatches 32; Indels 3; Gaps 1;

QY 1 QVQLQESGGGLVQAGGSLRLSCAASGSIISITWYRQAPGKERELVARMSSDGTTSY 60
DB 1 EVKLQESGGLVQPGGSLRLSCLASGTFSSNGSWVRQDPGGLQWADISSGGQTYA 60

QY 61 DSMKGRFTISRDNKNTVYLHNNLKPEDTAVYCYCKALISSYDGSWMDYGGQGTQVTVSS 120
DB 61 DAVKGRFISRDNAKNTLYLQMEDLRVEDTAVYCY--ATEGDIETPRYFGQGTIVTVSS 117

RESULT 12
HV37 MOUSE
ID HV37 MOUSE STANDARD; PRT; 119 AA.
AC P01807;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ig heavy chain V region X44.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE.
RX MEDLINE=79223895; PubMed=111245;
RA Rao D.N.; Rudikoff S.; Krutzsch H.; Potter M.;
RT "Structural evidence for independent joining region gene in
RT immunoglobulin heavy chains from anti-galactan myeloma proteins and
RT its potential role in generating diversity in
RT complementarity-determining regions."
RL Proc. Natl. Acad. Sci. U.S.A. 76:2890-2894(1979).
CC -!- MISCELLANEOUS: THIS CHAIN WAS ISOLATED FROM AN IGA MYELOMA PROTEIN
CC THAT BINDS GALACTAN.
CC -!- SIMILARITY: Contains 1 immunoglobulin-like domain.
DR PIR; A02077; AVMSX4.
DR HSSP; P01810; 2FB3.

DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003596; Ig_v.
DR Pfam; PF00047; Ig; 1.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS50835; IG LIKE; 1.
KW Immunoglobulin V region.
FT DOMAIN 1 116
FT NON_TER 117
FT NCBI_TaxID=9615;
SQ SEQUENCE 117 AA; 12703 MW; FCE0730950A84B35 CRC64;

Query Match 47.4%; Score 319.5; DB 1; Length 117;
Best Local Similarity 58.3%; Pred. No. 2e-24;
Matches 70; Conservative 15; Mismatches 32; Indels 3; Gaps 1;

QY 1 QVQLQESGGGLVQAGGSLRLSCLASGSIISITWYRQAPGKERELVARMSSDGTTSY 60
DB 1 EVKLQESGGLVQPGGSLRLSCLASGTFSSNGSWVRQDPGGLQWADISSGGQTYA 60

QY 61 DSMKGRFTISRDNKNTVYLHNNLKPEDTAVYCYCKALISSYDGSWMDYGGQGTQVTVSS 120
DB 61 DAVKGRFISRDNAKNTLYLQMEDLRVEDTAVYCY--ATEGDIETPRYFGQGTIVTVSS 117

RESULT 13
HV3B HUMAN
ID HV3B HUMAN STANDARD; PRT; 114 AA.
AC P01763;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ig heavy chain V-III region WEA.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
CX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE.
RX MEDLINE=83273707; PubMed=6410398;
RA Geni F.; Frangione B.;
RT "Amino acid sequence of the Fv region of a human monoclonal IgM
RT (protein WEA) with antibody activity against 3,4-pyruvylated
RT galactose in Klebsiella polysaccharides K30 and K33."
RL Proc. Natl. Acad. Sci. U.S.A. 80:4837-4841(1983).
CC -!- MISCELLANEOUS: THIS CHAIN WAS OBTAINED FROM A MONOCLONAL ANTIBODY
CC AGAINST 3,4-PYRUVYLATED GALACTOSE AND ISOLATED FROM A PATIENT WITH
CC WALDENSTROM'S MACROGLOBULINEMIA.
CC -!- SIMILARITY: Contains 1 immunoglobulin-like domain.
DR PIR; A02046; M3HWE.
DR HSSP; P01772; 2FB4.
DR GO; GO:0005576; C:extracellular; NAS.
DR GO; GO:0003823; F:antigen binding; NAS.
DR GO; GO:0006955; P:immune response; NAS.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003596; Ig_v.
DR Pfam; PF00047; Ig; 1.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS50835; IG LIKE; 1.
KW Immunoglobulin V region; Pyridolone carboxylic acid.
FT DOMAIN 1 112
FT MOD_RES 1 1 PYRROLIDONE CARBOXYLIC ACID.
FT NON_TER 114
FT NCBI_TaxID=9606;
SQ SEQUENCE 114 AA; 12256 MW; D88294FB418A07B7 CRC64;

Query Match 47.2%; Score 318; DB 1; Length 114;
Best Local Similarity 56.2%; Pred. No. 2.6e-24;
Matches 68; Conservative 17; Mismatches 28; Indels 8; Gaps 2;

QY 1 QVQLQESGGGLVQAGGSLRLSCLASGSIISITWYRQAPGKERELVARMSSDGTTSY 59
DB 1 EVKLQESGGLVQPGGSLRLSCLASGTFSSNGSWVRQDPGGLQWADISSGGQTYA 59
```

```
Db 1 QVQLVDSGGLVPEPGSLRLSCASGFTFSANDMNWVRQAPGKGLWLSFIQSGSTIYY 60
QY 60 TDSMKGRFTISRDNKNTVYLNHNKLPEDTAVYCKALISSYDGSWNDYGGQGTQVTVS 119
Db 61 ADSVKGRFTISRNBKSNLSYQLQMSLRASDPTAVYCAR-----GWLLNWGGTLTVTS 113
QY 120 S 120
Db 114 S 114

RESULT 14
HV38 MOUSE STANDARD; PRT; 119 AA.
AC P01806;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ig heavy chain V region T601.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_TaxID=10090;
RX [1]
RP SEQUENCE.
RX MEDLINE=79223895; PubMed=111245;
RA Rao D.N., Rudikoff S., Kruttsch H., Potter M.;
RT "Structural evidence for independent joining region gene in
RT immunoglobulin heavy chains from anti-galactan myeloma proteins and
RT its potential role in generating diversity in
RT complementarity-determining regions."
RL Proc. Natl. Acad. Sci. U.S.A. 76:2890-2894(1979).
CC -!- MISCELLANEOUS: THIS CHAIN WAS ISOLATED FROM AN IGA MYELOMA PROTEIN
CC THAT BINDS GALACTAN.
CC -!- SIMILARITY: Contains 1 immunoglobulin-like domain.
DR PIR; A02078; AVMSt6.
DR HSSP; P01810; 2FBU.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003596; Ig_V.
DR Pfam; PF00047; Ig; 1.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS0835; IG-LIKE; 1.
KW Immunoglobulin V region.
FT DOMAIN 1 112 IG-LIKE.
FT NON TER 119 119
SQ SEQUENCE 119 AA; 13169 MW; BC38CC84E6EA00E8 CRC64;

Query Match 47.0%; Score 317; DB 1; Length 119;
Best Local Similarity 57.4%; Pred. No. 3.5e-24;
Matches 70; Conservative 17; Mismatches 31; Indels 4; Gaps 3;

QY 1 QVQLVDSGGLVQAGGSLRLSCAASGSRISIMTWYRQAPGKERELVARMSSDGTGTT-SY 59
Db 1 EVKLLSEGGGLVQPGGSLKSLCAASGDFSRYSWMSWRQAPGKGLWIGWEINPDSTINY 60
QY 60 TDSMKGRFTISRDNKNTVYLNHNKLPEDTAVYCKALISSYDGSWNDYGGQGTQVTVS 119
Db 61 TPSLKDKFIISRDNAKNTLYLQMSKVRSDTALYCARL--GYIG-YFDVWVGAGTTVTS 117
QY 120 SE 121
Db 118 SE 119

RESULT 15
HV3U HUMAN STANDARD; PRT; 120 AA.
AC P01782;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ig heavy chain V-III region DOB.
OS Homo sapiens (Human).
```

```
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC NCBI_TaxID=9606;
RX [1]
RP SEQUENCE.
RX MEDLINE=80020921; PubMed=114209;
RA Steiner L.A., Garcia Pardo A., Margolies M.N.;
RT "Amino acid sequence of the heavy-chain variable region of the
RT crystallizable human myeloma protein Dob."
RL Biochemistry 18:4068-4080(1979).
RX [2]
RP CRYSTALLIZATION.
RX MEDLINE=80020920; PubMed=114208;
RA Steiner L.A., Lopes A.D.;
RT "The crystallizable human myeloma protein Dob has a hinge-region
RT deletion."
RL Biochemistry 18:4054-4067(1979).
CC -!- MISCELLANEOUS: THIS GAMMA-1 MYELOMA PROTEIN HAS A DELETION IN THE
CC HINGE REGION. THERE ARE NO LIGHT-HEAVY OR INTER-HEAVY CHAIN
CC DISULFIDE BONDS.
CC -!- SIMILARITY: Contains 1 immunoglobulin-like domain.
DR PIR; A90431; GIHUDB.
DR HSSP; P01772; 2F34.
DR GO; GO:0005576; C:extracellular; NAS.
DR GO; GO:0003823; F:antigen binding; NAS.
DR GO; GO:0006955; P:immune response; NAS.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003596; Ig_V.
DR Pfam; PF00047; Ig; 1.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS0835; IG-LIKE; 1.
KW Immunoglobulin V region.
FT DOMAIN 1 112 IG-LIKE.
FT NON TER 120 120
SQ SEQUENCE 120 AA; 13440 MW; 880DDE307C4B2627 CRC64;
```

Query Match 46.7%; Score 315; DB 1; Length 120;  
Best Local Similarity 58.2%; Pred. No. 5.5e-24;  
Matches 71; Conservative 14; Mismatches 33; Indels 4; Gaps 3;

```
QY 1 QVQLVDSGGLVQAGGSLRLSCAASGSRISIMTWYRQAPGKERELVARMSSDGTTSY 59
Db 1 EVQLVESGGDLVQPGFSLRLSCAASGFRFHYNNHNLROGPGKPEWVSTITWNGGSVLY 60
QY 60 TDSMKGRFTISRDNKNTVYLNHNKLPEDTAVYCY-KALISSYDGSWNDYGGQGTQVTV 118
Db 61 ADSVKGRFAISRDNKNTLYLQNLRLPEDTAPYCAKGYI--WNGNWFDSWGQGLTVTV 118
QY 119 SS 120
Db 119 SS 120
```

Search completed: October 5, 2004, 08:01:57  
Job time : 12.5463 secs

This Page Blank (uspic)

GenCore version 5.1.1.6  
 Copyright (c) 1993 - 2004 Compugen Ltd.  
 OM protein - protein search, using sw model  
 Run on: October 5, 2004, 07:34:11 ; Search time 68.2824 Seconds  
 (without alignments)  
 596.081 Million cell updates/sec

Title: US-09-805-290A-25  
 Perfect score: 674  
 Sequence: 1 QVQLQESGGGLVQAGGSLRL.....GGQGTVTVSSSEPKTPKQP 129

Scoring table: BLOSUM62  
 Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL\_25.\*

- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mhc.\*
- 8: sp\_organelle.\*
- 9: sp\_phase.\*
- 10: sp\_plant.\*
- 11: sp\_rodent.\*
- 12: sp\_virus.\*
- 13: sp\_vertebrate.\*
- 14: sp\_unclassified.\*
- 15: sp\_rvirus.\*
- 16: sp\_bacteriap.\*
- 17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	361.5	53.6	597	Q96BB9	Q96bb9 homo sapien
2	351	52.1	118	Q9UL72	Q9ul72 homo sapien
3	341.5	50.7	480	Q91XE1	Q91xel mus musculus
4	341	50.6	487	Q99KA4	Q99ka4 mus musculus
5	341	50.6	613	Q8WU1	Q8wuk1 homo sapien
6	340.5	50.5	113	Q9UL90	Q9ul90 homo sapien
7	340.5	50.5	573	Q8WU38	Q8wu38 homo sapien
8	338	50.1	499	Q8N5K4	Q8n5k4 homo sapien
9	337.5	50.1	479	Q91WP5	Q91wp5 mus musculus
10	336.5	49.9	131	Q9UL88	Q9ul88 homo sapien
11	335.5	49.8	116	Q9UL93	Q9ul93 homo sapien
12	334.5	49.6	121	Q9UL71	Q9ul71 homo sapien
13	334.5	49.6	494	Q96K68	Q96k68 homo sapien
14	334	49.6	118	Q9UL91	Q9ul91 homo sapien
15	331.5	49.2	486	Q91Z07	Q91z07 mus musculus
16	330.5	49.0	119	Q920E7	Q920e7 mus musculus

17	330.5	49.0	147	4	Q9Y509	Q9y509 homo sapien
18	329	48.8	298	11	Q9QYF0	Q9qyf0 mus musculus
19	324.5	48.1	95	4	Q9ULB6	Q9ulb6 homo sapien
20	324	48.1	473	11	Q91Z05	Q91z05 mus musculus
21	322	47.8	122	4	Q9UL84	Q9ul84 homo sapien
22	322	47.8	493	4	Q8NCL6	Q8nc16 homo sapien
23	319	47.3	112	4	Q9HCC1	Q9hcc1 homo sapien
24	318.5	47.3	487	11	Q80Z17	Q80z17 mus musculus
25	318	47.2	437	11	Q9RIA4	Q9ria4 mus musculus
26	310	46.0	469	11	Q8R3V9	Q8r3v9 mus musculus
27	307	45.5	521	4	Q8N4Y9	Q8n4y9 homo sapien
28	300	44.5	470	4	Q7Z5W1	Q7z5w1 homo sapien
29	294	43.6	484	11	Q8VEA0	Q8vea0 mus musculus
30	286	42.4	479	11	Q7TMC4	Q7tnk4 mus musculus
31	284	42.1	482	4	Q7Z351	Q7z351 homo sapien
32	275	40.8	482	11	Q91X92	Q91x92 mus musculus
33	272	40.4	124	6	Q9N0W6	Q9n0w6 oryctolagus
34	270	40.1	124	6	Q9N0W4	Q9n0w4 oryctolagus
35	269.5	40.0	112	4	Q9UGP3	Q9ugp3 homo sapien
36	259	38.4	143	11	Q924P9	Q924p9 mus musculus
37	258	38.3	121	11	Q99NG4	Q99ng4 mus musculus
38	255.5	37.9	614	11	Q7TWT6	Q7tnk6 mus musculus
39	254.5	37.8	104	4	Q9UL87	Q9ul87 homo sapien
40	253.5	37.6	119	4	Q9UL73	Q9ul73 homo sapien
41	253.5	37.6	479	11	Q99M22	Q99m22 mus musculus
42	251	37.2	124	4	Q9UL92	Q9ul92 homo sapien
43	251	37.2	145	11	Q924Q9	Q924q9 mus musculus
44	251	37.2	145	11	Q924R1	Q924r1 mus musculus
45	250.5	37.2	146	11	Q924Q3	Q924q3 mus musculus

## ALIGNMENTS

## RESULT 1

Q96BB9 PRELIMINARY; PRT; 597 AA.  
 AC Q96BB9;  
 DT 01-DEC-2001 (Tremblrel. 19, Created)  
 DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)  
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)  
 DE Hypothetical protein.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=B-cell;  
 RA Strausberg R.;  
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC015760; AAH15760.1; -  
 DR InterPro; IPR007110; Ig-like.  
 DR InterPro; IPR003006; Ig\_MHC.  
 DR InterPro; IPR003596; Ig\_V.  
 DR Pfam; PF00047; Ig; 5.  
 DR SMART; SM00406; IG; 1.  
 DR PROSITE; PS00835; IG LIKE; 5.  
 DR PROSITE; PS00290; IG\_MHC; 3.  
 KW Hypothetical protein.  
 SQ SEQUENCE 597 AA; 65039 MW; 4FCA3AD8CE263D9 CRC64;

Query Match 53.6%; Score 361.5; DB 4; Length 597;

Best Local Similarity 53.2%; Pred. No. 3.4e-27;

Matches 79; Conservative 11; Mismatches 30; Indels 5; Gaps 2;

Qy 1 QVQLQESGGGLVQAGGSLRLSCAASGSIIRISITWVROAPGKERELVARMS-SDGTTSY 59

Db 20 EVQLLESQGLVPGGSLRLSCAASGFSFSSYANWVRQAPGKLEWVSISGGSTYY 79

Qy 60 TDSMKGFTTIRDNKNTVYLHVNLIKPEDTAVYCKALISSYDGSWN----DYGGQGTQ 115

Db 80 ADSVKGRFTTIRDNRSRDTLYLQWNSLRAEDTAVVYCAKDPGYSASGNTVREDYWGQGT 139

```

QY 116 VTSS 120
DB 140 VTSS 144

RESULT 2
Q9UL72 PRELIMINARY; PRT; 118 AA.
AC Q9UL72;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Myosin-reactive immunoglobulin heavy chain variable region
DE (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98277139; PubMed=9614934;
RA Wu X., Liu B., Van der Merwe P.L., Kallis N.N., Berney S.M.,
RA "Myosin-reactive autoantibodies in rheumatic carditis and normal
RA fetus.";
RL Clin. Immunol. Immunopathol. 87:184-192(1998).
DR EMBL; AF035042; AAD56278.1; -.
DR PIR; S21205; S21205.
DR HSSP; P01772; 2PB4.
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003596; IG_v.
DR Pfam; PF00047; IG; 1.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS00835; IG_LIKE; 1.
FT NON_TER 1
FT NON_TER 118
SQ SEQUENCE 118 AA; 12872 MW; B4D1A5944B2D5CCA CRC64;

Query Match 52.1%; Score 351; DB 4; Length 118;
Best Local Similarity 62.5%; Pred. No. 4.5e-27;
Matches 75; Conservative 11; Mismatches 32; Indels 2; Gaps 1;

QY 1 QVQLQESGGGLVQAGGSLRLSCAASGSIISIMTWYRQAPGKERELVARMSDGTTSYT 60
DB 1 EVQLVESGGGLVQGGSLRLSCAASGFTVSSNYMWVRQPGKLGESVSYTSGGSSYYA 60

QY 61 DSMKGRFTISRDNKNTVYLHMNLPKEDTAVYYCKALISSYDGSWMDYGGQGTQTVSS 120
DB 61 DSVKGRFTISRDNKNTLYLQNSLRAEDTAFYYCAR--DRFGFLFDYWGQGLTVTVSS 118

RESULT 3
Q9IXE1 PRELIMINARY; PRT; 480 AA.
AC Q9IXE1;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Hypothetical protein (fragment).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=Colon;
RA Strausberg R.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBSJ databases.
DR EMBL; BC010798; AAH10798.1; -.
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003006; IG_MHC.
DR InterPro; IPR003596; IG_v.

Query Match 50.6%; Score 341; DB 11; Length 487;
Best Local Similarity 60.8%; Pred. No. 2.8e-25;
Matches 76; Conservative 14; Mismatches 31; Indels 4; Gaps 3;

QY 1 QVQLQESGGGLVQAGGSLRLSCAASGSIISIMTWYRQAPGKERELVARMSDGT-TSY 59
DB 20 EVQLVESGGGLVQGGSLRLSCAASGFTFSSYAMSVVRQTPKRLWVATISDGSYYT 79

QY 60 TDSMKGRFTISRDNKNTVYLHMNLPKEDTAVYYCKALI--SSYDGSWMDYGGQGTQV 116
DB 80 PDNVKGRFTISRDNKNNLYLQMSHLKSEDITAMYYCARDMGSPYGGYSRFDYWGQGTI 139

RESULT 4
Q99XA4 PRELIMINARY; PRT; 487 AA.
AC Q99XA4;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Hypothetical protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (MAR-2001) to the EMBL/GenBank/DBSJ databases.
DR EMBL; BC004786; AAH04786.1; -.
DR HSSP; P01810; 2PBC.
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003006; IG_MHC.
DR InterPro; IPR003596; IG_v.
DR Pfam; PF00047; IG; 4.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS00835; IG_LIKE; 4.
DR PROSITE; PS00290; IG_MHC; 2.
KW Hypothetical protein.
SQ SEQUENCE 487 AA; 52554 MW; 7DC8E96DEB333077B CRC64;

Query Match 50.6%; Score 341; DB 11; Length 487;
Best Local Similarity 60.8%; Pred. No. 2.8e-25;
Matches 76; Conservative 14; Mismatches 31; Indels 4; Gaps 3;

QY 1 QVQLQESGGGLVQAGGSLRLSCAASGSIISIMTWYRQAPGKERELVARMSDGT-TSY 59
DB 20 EVQLVESGGGLVQGGSLRLSCAASGFTFSSYAMSVVRQTPKRLWVATISDGSYYT 79

QY 60 TDSMKGRFTISRDNKNTVYLHMNLPKEDTAVYYCKALI--SSYDGSWMDYGGQGTQV 116
DB 80 PDNVKGRFTISRDNKNNLYLQMSHLKSEDITAMYYCARDMGSPYGGYSRFDYWGQGTI 139

RESULT 5
Q8WUK1 PRELIMINARY; PRT; 613 AA.
ID Q8WUK1

```



Q8WUK1;  
DT 01-MAR-2002 (Tremblrel. 20, Created)  
DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)  
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)  
DE Hypothetical protein.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Tonsil;  
RA Strausberg R.;  
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BC020240; AAH20240.1; -;  
DR PIR; P0120; P0120.  
DR PIR; S15590; S15590.  
DR InterPro; IPR007110; Ig-like.  
DR InterPro; IPR003006; Ig\_MHC.  
DR InterPro; IPR003596; Ig\_V.  
DR Pfam; PF00047; Ig; 1.  
DR SMART; SM00406; IGV; 1.  
DR PROSITE; PS0835; IG\_LIKE; 5.  
DR PROSITE; PS00290; IG\_MHC; 3.  
KW Hypothetical protein.  
SQ SEQUENCE 613 AA; 67296 MW; 50C7F5950671B315 CRC64;  
Query Match 50.6%; Score 341; DB 4; Length 613;  
Best Local Similarity 60.6%; Pred. No. 3.8e-25;  
Matches 77; Conservative 9; Mismatches 27; Indels 14; Gaps 3;  
QY 1 QVQLQESGGGLVQAGGSLRLSCAASGSIISITMTWYRQAPGKERELVARMSSDGTTS-Y 59  
20 QVQLVESGGGVQPGGSLRLSCAASGFTFSYGMHWVRQAPGKLEWVAFIRYDGSNKYY 79  
Db  
QY 60 TDSMKGRFTISRDNKNTVYLMNKLKPEDTAVYCKALISSYDGSW-----DYGGQG 113  
80 ADSVKGRFTISRDNKNTLYLQMSLRADTAVYICAK-----DWSBGEVETDIWGGQ 132  
114 TQVTSS 120  
133 TMTVSS 139  
Db  
RESULT 6  
Q9UL90  
ID Q9UL90 PRELIMINARY; PRT; 113 AA.  
AC Q9UL90;  
DT 01-MAY-2000 (Tremblrel. 13, Created)  
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)  
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)  
DE Myosin-reactive immunoglobulin heavy chain variable region (Fragment)  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE-98277139; PubMed-9614934;  
RA Wu X., Liu B., Van der Merwe P.L., Kalis N.N., Berney S.M.,  
RA Young D.C.;  
RT "Myosin-reactive autoantibodies in rheumatic carditis and normal fetus."  
RT Clin. Immunol. Immunopathol. 87:184-192(1998).  
RL EMBL; AF035024; AAD56260.1; -;  
DR HSSP; P01772; 2FB4.  
DR InterPro; IPR007110; Ig-like.  
DR InterPro; IPR003596; Ig\_V.  
DR Pfam; PF00047; Ig; 1.  
DR SMART; SM00406; IGV; 1.  
DR PROSITE; PS0835; IG\_LIKE; 1.  
NON\_TER 1 1

FT NON\_TER 113 113  
SQ SEQUENCE 113 AA; 12437 MW; ED57FDD19086D07F CRC64;  
Query Match 50.5%; Score 340.5; DB 4; Length 113;  
Best Local Similarity 62.0%; Pred. No. 4.7e-26;  
Matches 75; Conservative 11; Mismatches 26; Indels 9; Gaps 2;  
QY 1 QVQLQESGGGLVQAGGSLRLSCAASGSIISITMTWYRQAPGKERELVARMSSDGTTS-Y 59  
1 EVQLVESGGGVQPGGSLRLSCAASGFTFSYGMHWVRQAPGKLEWVAFIRYDGSNKYY 60  
Db  
QY 60 TDSMKGRFTISRDNKNTVYLMNKLKPEDTAVYCKALISSYDGSWNDYGGQGTQVTS 119  
61 ADSVKGRFTISRDNKNTLYLQMSLRADTAVYICAKL-----NYWGQGLTQVTS 112  
QY 120 S 120  
Db 113 S 113  
RESULT 7  
Q8WU38  
ID Q8WU38 PRELIMINARY; PRT; 573 AA.  
AC Q8WU38;  
DT 01-MAR-2002 (Tremblrel. 20, Created)  
DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)  
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)  
DE Hypothetical protein.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Tonsil;  
RA Strausberg R.;  
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BC021276; AAH21276.1; -;  
DR PIR; S21205; S21205.  
DR InterPro; IPR007110; Ig-like.  
DR InterPro; IPR003006; Ig\_MHC.  
DR InterPro; IPR003596; Ig\_V.  
DR Pfam; PF00047; Ig; 4.  
DR SMART; SM00406; IGV; 1.  
DR PROSITE; PS0835; IG\_LIKE; 4.  
DR PROSITE; PS00290; IG\_MHC; 2.  
KW Hypothetical protein.  
SQ SEQUENCE 573 AA; 62967 MW; FD072344033AC530 CRC64;  
Query Match 50.5%; Score 340.5; DB 4; Length 573;  
Best Local Similarity 59.7%; Pred. No. 3.9e-25;  
Matches 80; Conservative 12; Mismatches 37; Indels 5; Gaps 4;  
QY 1 QVQLQESGGGLVQAGGSLRLSCAASGSIISITMTWYRQAPGKERELVARMSSDGTTSY 59  
20 EVQLVESGGGLVQPGGSLRLSCAASGFTFDYAMHWVRQAPGKLEWVSGISWNSGIGY 79  
Db  
QY 60 TDSMKGRFTISRDNKNTVYLMNKLKPEDTAVYCKALISSYDGSW--NDYGGQGTQV 116  
80 ADSVKGRFTISRDNKNTLYLQMSLRADTAVYICAKHGSGSYIGYYGMDVWGQGITV 139  
QY 117 TVSSEP-KTPKQP 129  
Db 140 TVSSAPTAPDVFP 153  
RESULT 8  
Q8NSK4  
ID Q8NSK4 PRELIMINARY; PRT; 499 AA.  
AC Q8NSK4;  
DT 01-OCT-2002 (Tremblrel. 22, Created)  
DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)  
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)



OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 RN NCBI\_TaxID=9606;  
 RX [1]  
 RP SEQUENCE FROM N.A.  
 RZ MEDLINE=98277139; PubMed=9614934;  
 RA Wu X., Liu B., Van der Merwe P.L., Kalis N.N., Berney S.M.,  
 RT "Myosin-reactive autoantibodies in rheumatic carditis and normal  
 fetus";  
 RL Clin. Immunol. Immunopathol. 87:184-192(1998).  
 DR EMBL: AF035021; AAD56257.1; -.  
 DR PIR: P0120; P0120.  
 DR HSSP: P01772; 2FB4.  
 DR InterPro: IPR007110; Ig-like.  
 DR InterPro: IPR003596; IG\_V.  
 DR Pfam: PF00047; IG; 1.  
 DR SMART: SM00406; IG; 1.  
 DR PROSITE: PS00835; IG\_LIKE; 1.  
 FT NON\_TER 1 1  
 FT TER 116 1  
 SQ SEQUENCE 116 AA; 12434 MW; 0DA0348154DD6061 CRC64;

Query Match 49.8%; Score 335.5; DB 4; Length 116;  
 Best Local Similarity 62.5%; Pred. No. 1.5e-25;  
 Matches 75; Conservative 10; Mismatches 30; Indels 5; Gaps 2;

QY 2 VQLQESGGGLVQAGGSLRLSCLASGSIRISIMTWYRQAPGKERELVARMSSDGTTS-YT 60  
 Db 1 VQLVESGGGVVQPGGSLRLSCLASGSIRISIMTWYRQAPGKERELVARMSSDGTTS 60  
 QY 61 DSMKGRFTISRDNKNTVILHMNLKPEPTAVYVC-KALISS-YDGSWYDYGSGGTQVT 120  
 Db 61 DSVKGRFTISRDNKNTVILHMNLKPEPTAVYVC-KALISS-YDGSWYDYGSGGTQVT 116

## RESULT 12

Q9UL71 PRELIMINARY; PRT; 121 AA.  
 ID Q9UL71  
 AC Q9UL71  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Myosin-reactive immunoglobulin heavy chain variable region  
 DE (Fragment).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 RN NCBI\_TaxID=9606;  
 RX [1]  
 RP SEQUENCE FROM N.A.  
 RZ MEDLINE=98277139; PubMed=9614934;  
 RA Wu X., Liu B., Van der Merwe P.L., Kalis N.N., Berney S.M.,  
 RT "Myosin-reactive autoantibodies in rheumatic carditis and normal  
 fetus";  
 RL Clin. Immunol. Immunopathol. 87:184-192(1998).  
 DR EMBL: AF035043; AAD56279.1; -.  
 DR HSSP: P01772; 2FB4.  
 DR InterPro: IPR007110; Ig-like.  
 DR InterPro: IPR003596; IG\_V.  
 DR Pfam: PF00047; IG; 1.  
 DR SMART: SM00406; IG; 1.  
 DR PROSITE: PS00835; IG\_LIKE; 1.  
 FT NON\_TER 1 1  
 FT TER 121 1  
 SQ SEQUENCE 121 AA; 13154 MW; 2F045CCFA5D50736 CRC64;

Query Match 49.6%; Score 334.5; DB 4; Length 121;  
 Best Local Similarity 61.0%; Pred. No. 2e-25;  
 Matches 75; Conservative 15; Mismatches 28; Indels 5; Gaps 4;

QY 1 QVQLQESGGGLVQAGGSLRLSCLASGSIRISIMTWYRQAPGKERELVARMSSD-GTTSY 59  
 Db 1 EVQLVESGGGVVQPGGSLRLSCLASGSIRISIMTWYRQAPGKERELVARMSSD-GTTSY 60  
 QY 60 TDSMKGRFTISRDNKNTVILHMNLKPEPTAVYVC-KALISS-YDGSWYDYGSGGTQVT 117  
 Db 61 ADSVKGRTISRDNKNTVILHMNLKPEPTAVYVC-KALISS-YDGSWYDYGSGGTQVT 118  
 QY 118 VSS 120  
 Db 119 VSS 121

## RESULT 13

Q96K68 PRELIMINARY; PRT; 494 AA.  
 ID Q96K68  
 AC Q96K68;  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Hypothetical protein FLJ1473.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 RN NCBI\_TaxID=9606;  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Mammary gland;  
 RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,  
 RA Nishikawa T., Nagai K., Sugano S., Shitatori A., Sudo H.,  
 RA Watanabe M., Hosoi T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,  
 RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,  
 RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,  
 RA Yamamoto J., Wakamatsu A., Nakamura Y., Negahari K., Masuho Y.,  
 RA Kinomiyama K., Iwayanagi T.;  
 RT "NEDO human cDNA sequencing project";  
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AK027379; BAB55072.1; -.  
 DR PIR: S21205; S21205.  
 DR InterPro: IPR007110; Ig-like.  
 DR InterPro: IPR003006; IG\_MHC.  
 DR InterPro: IPR003596; IG\_V.  
 DR Pfam: PF00047; IG; 4.  
 DR SMART: SM00406; IG; 1.  
 DR PROSITE: PS00835; IG\_LIKE; 4.  
 DR PROSITE: PS00290; IG\_MHC; 1.  
 KW Hypothetical protein  
 SQ SEQUENCE 494 AA; 53088 MW; 9A1D7AE85AE34C0E CRC64;

Query Match 49.6%; Score 334.5; DB 4; Length 494;  
 Best Local Similarity 57.2%; Pred. No. 1.3e-24;  
 Matches 79; Conservative 13; Mismatches 31; Indels 15; Gaps 5;

QY 1 QVQLQESGGGLVQAGGSLRLSCLASGSIRISIMTWYRQAPGKERELVARMSS-DGTTSY 59  
 Db 20 EVQLVESGGGVVQPGGSLRLSCLASGSIRISIMTWYRQAPGKERELVARMSS-DGTTSY 79  
 QY 60 TDSMKGRFTISRDNKNTVILHMNLKPEPTAVYVC-KALISSYDGS-----WNYDGG 112  
 Db 80 RDSVKGRTISRDNKNTVILHMNLKPEPTAVYVCAR--DSCNGAICYGSPW----GQ 133  
 QY 113 GTQTVSS-EKTPKQP 129  
 Db 134 GTLVTSASPTSPKVP 151

## RESULT 14

Q9UL91 PRELIMINARY; PRT; 118 AA.  
 ID Q9UL91  
 AC Q9UL91;  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)





XX The patent discloses antibodies or their fragments comprising a heavy  
 CC chain variable domain (VHH) derived from an immunoglobulin naturally  
 CC devoid of light chains specific for inhibiting human dietary enzymes. The  
 CC antibodies of the invention are useful for the preparation of medicaments  
 CC or food for inhibiting the activity of one or more human dietary enzymes  
 CC especially human pancreatic lipase (HPL) or human gastric lipase (HGL)  
 CC which are useful for the cosmetic control of body weight of human beings.  
 CC The present peptide sequence is HPL inhibiting VHH fragment, HPL #22 from  
 CC llama (camelid) species  
 XX  
 SQ Sequence 129 AA;

Query Match 100.0%; Score 674; DB 4; Length 129;  
 Best Local Similarity 100.0%; Pred. No. 3.7e-55;  
 Matches 129; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QVQLQESGGGLVQAGGSLRLSCAASGSTRISINTWYRQAPGKERELVARMSSDGTTSYT 60  
 |||||  
 DB 1 QVQLQESGGGLVQAGGSLRLSCAASGSTRISINTWYRQAPGKERELVARMSSDGTTSYT 60  
 |||||  
 QY 61 DSMKGRFTISRDNKNTVYLHMNNLKPEDTAVYYCKALISSYDGSWNDYGGQGTQVTSS 120  
 |||||  
 DB 61 DSMKGRFTISRDNKNTVYLHMNNLKPEDTAVYYCKALISSYDGSWNDYGGQGTQVTSS 120  
 |||||  
 QY 121 EPKTPKPOP 129  
 |||||  
 DB 121 EPKTPKPOP 129  
 |||||

RESULT 2  
 AAE10553  
 ID AAE10553 standard; peptide; 129 AA.

XX AAE10553;  
 AC AAE10553;  
 DT 10-DEC-2001 (first entry)  
 XX HPL inhibiting VHH fragment, HPL #13 from llama species.  
 XX Llama antibody; camelid; anorectic; heavy chain variable domain; VHH;  
 KW human dietary enzyme inhibitor; medicament; human pancreatic lipase; HPL;  
 KW food; human gastric lipase; HGL; cosmetic control; body weight.  
 XX  
 OS Lama sp.  
 XX Location/Qualifiers  
 FH Key 31..35  
 FT Region /label= CDR1  
 FT /note= "Complementarity determining region 1"  
 FT Region 50..64  
 FT /label= CDR2  
 FT /note= "Complementarity determining region 2"  
 FT Region 98..109  
 FT /label= CDR3  
 FT /note= "Complementarity determining region 3"  
 XX  
 PN EP1134231-A1.  
 XX 19-SEP-2001.  
 XX 20-FEB-2001; 2001EP-00200703.  
 XX 14-MAR-2000; 2000EP-00200930.  
 XX (UNIL ) UNILEVER NV.  
 XX (UNIL ) UNILEVER PLC.  
 XX Bezemer S, Van De Burg M, De Haard JMW, Tareilus E;  
 PI WPI; 2001-572718/65.  
 XX New antibody or its fragments for inhibiting human dietary enzymes,

PT useful for cosmetic control of body weight of human beings, comprises  
 PT heavy chain variable domain derived from immunoglobulin naturally devoid  
 PT of light chains.  
 XX Example 2; Page 9; 37pp; English.  
 PS The patent discloses antibodies or their fragments comprising a heavy  
 CC chain variable domain (VHH) derived from an immunoglobulin naturally  
 CC devoid of light chains specific for inhibiting human dietary enzymes. The  
 CC antibodies of the invention are useful for the preparation of medicaments  
 CC or food for inhibiting the activity of one or more human dietary enzymes  
 CC especially human pancreatic lipase (HPL) or human gastric lipase (HGL)  
 CC which are useful for the cosmetic control of body weight of human beings.  
 CC The present peptide sequence is HPL inhibiting VHH fragment, HPL #13 from  
 CC llama (camelid) species  
 XX  
 SQ Sequence 129 AA;

Query Match 79.7%; Score 537; DB 4; Length 129;  
 Best Local Similarity 79.8%; Pred. No. 2.3e-42;  
 Matches 103; Conservative 11; Mismatches 15; Indels 0; Gaps 0;

QY 1 QVQLQESGGGLVQAGGSLRLSCAASGSTRISINTWYRQAPGKERELVARMSSDGTTSYT 60  
 |||||  
 DB 1 QVQLQESGGGLVQAGGSLRLSCAASGSTRISINTWYRQAPGKERELVARMSSDGTTSYT 60  
 |||||  
 QY 61 DSMKGRFTISRDNKNTVYLHMNNLKPEDTAVYYCKALISSYDGSWNDYGGQGTQVTSS 120  
 |||||  
 DB 61 DSMKGRFTISRDNKNTVYLHMNNLKPEDTAVYYCKALISSYDGSWNDYGGQGTQVTSS 120  
 |||||  
 QY 121 EPKTPKPOP 129  
 |||||  
 DB 121 EPKTPKPOP 129  
 |||||

RESULT 3  
 AAE10555  
 ID AAE10555 standard; peptide; 131 AA.

XX AAE10555;  
 AC AAE10555;  
 DT 10-DEC-2001 (first entry)  
 XX HPL inhibiting VHH fragment, HPL #15 from llama species.  
 XX Llama antibody; camelid; anorectic; heavy chain variable domain; VHH;  
 KW human dietary enzyme inhibitor; medicament; human pancreatic lipase; HPL;  
 KW food; human gastric lipase; HGL; cosmetic control; body weight.  
 XX  
 OS Lama sp.  
 XX Location/Qualifiers  
 FH Key 31..35  
 FT Region /label= CDR1  
 FT /note= "Complementarity determining region 1"  
 FT Region 50..64  
 FT /label= CDR2  
 FT /note= "Complementarity determining region 2"  
 FT Region 98..111  
 FT /label= CDR3  
 FT /note= "Complementarity determining region 3"  
 XX  
 PN EP1134231-A1.  
 XX 19-SEP-2001.  
 XX 20-FEB-2001; 2001EP-00200703.  
 XX 14-MAR-2000; 2000EP-00200930.  
 XX (UNIL ) UNILEVER NV.  
 XX (UNIL ) UNILEVER PLC.  
 XX Bezemer S, Van De Burg M, De Haard JMW, Tareilus E;  
 PI WPI; 2001-572718/65.  
 XX New antibody or its fragments for inhibiting human dietary enzymes,

PI Bezemer S, Van De Burg M, De Haard JJW, Tareilus E;  
XX WPI; 2001-572718/65.  
XX  
XX New antibody or its fragments for inhibiting human dietary enzymes,  
PT useful for cosmetic control of body weight of human beings, comprises  
PT heavy chain variable domain derived from immunoglobulin naturally devoid  
PT of light chains.  
XX  
XX Example 2; Page 9; 37pp; English.  
XX  
XX The patent discloses antibodies or their fragments comprising a heavy  
XX chain variable domain (VHH) derived from an immunoglobulin naturally  
CC devoid of light chains specific for inhibiting human dietary enzymes. The  
CC antibodies of the invention are useful for the preparation of medicaments  
CC or food for inhibiting the activity of one or more human dietary enzymes  
CC especially human pancreatic lipase (HPL) or human gastric lipase (HGL)  
CC which are useful for the cosmetic control of body weight of human beings.  
CC The present peptide sequence is HPL inhibiting VHH fragment, HPL #15 from  
XX llama (camelid) species  
XX  
XX Sequence 131 AA;  
Query Match 75.4%; Score 508; DB 4; Length 131;  
Best Local Similarity 77.1%; Pred. No. 1.2e-39;  
Matches 101; Conservative 11; Mismatches 17; Indels 2; Gaps 1;  
QY 1 QVQLQSGGGLVQAGGSLRLSCAASGSIIRISIMTWYQAPGKERELVARMSSDGTTSYT 60  
DB 1 QVQLQSGGGLVQAGGSLRLSCAASGSIIRISIMTWYQAPGKERELVARMSSDGTTSYA 60  
QY 61 DSMKGRFTISRDNKNTVYLMHNNLKPEDTAVYICKALISSYDGSWMDYGGQGTQTVV 118  
DB 61 DSLKGRFTISRDNKNTVYLMHNNLKPEDTAVYICNAHITPAGSNVYVYGYWGHGKTVV 120  
QY 119 SSEPKTPKPQ 129  
DB 121 SSEPKTPKPQ 131  
RESULT 4  
AAE10551  
ID AAE10551 standard; peptide; 129 AA.  
XX  
XX AAE10551;  
XX  
XX 10-DEC-2001 (first entry)  
XX  
XX HPL inhibiting VHH fragment, HPL #11 from llama species.  
XX  
XX Llama antibody; camelid; anorectic; heavy chain variable domain; VHH;  
KW human dietary enzyme inhibitor; medicament; human pancreatic lipase; HPL;  
KW food; human gastric lipase; HGL; cosmetic control; body weight.  
XX  
XX Lama sp.  
XX  
XX Location/Qualifiers  
FH Key 31..35  
FT Region /label= CDR1  
FT /note= "Complementarity determining region 1"  
FT Region 50..64  
FT /label= CDR2  
FT /note= "Complementarity determining region 2"  
FT Region 98..109  
FT /label= CDR3  
FT /note= "Complementarity determining region 3"  
XX  
XX EPI134231-A1.  
XX  
XX 19-SEP-2001.  
XX  
XX 20-FEB-2001; 2001EP-00200703.  
XX  
XX

14-MAR-2000; 2000EP-00200930.  
XX (UNIL ) UNILEVER NV.  
PA (UNIL ) UNILEVER PLC.  
XX  
XX Bezemer S, Van De Burg M, De Haard JJW, Tareilus E;  
XX WPI; 2001-572718/65.  
XX  
XX New antibody or its fragments for inhibiting human dietary enzymes,  
PT useful for cosmetic control of body weight of human beings, comprises  
PT heavy chain variable domain derived from immunoglobulin naturally devoid  
PT of light chains.  
XX  
XX Example 2; Page 9; 37pp; English.  
XX  
XX The patent discloses antibodies or their fragments comprising a heavy  
XX chain variable domain (VHH) derived from an immunoglobulin naturally  
CC devoid of light chains specific for inhibiting human dietary enzymes. The  
CC antibodies of the invention are useful for the preparation of medicaments  
CC or food for inhibiting the activity of one or more human dietary enzymes  
CC especially human pancreatic lipase (HPL) or human gastric lipase (HGL)  
CC which are useful for the cosmetic control of body weight of human beings.  
CC The present peptide sequence is HPL inhibiting VHH fragment, HPL #11 from  
XX llama (camelid) species  
XX  
XX Sequence 129 AA;  
Query Match 74.3%; Score 501; DB 4; Length 129;  
Best Local Similarity 77.2%; Pred. No. 5.4e-39;  
Matches 98; Conservative 9; Mismatches 20; Indels 0; Gaps 0;  
QY 1 QVQLQSGGGLVQAGGSLRLSCAASGSIIRISIMTWYQAPGKERELVARMSSDGTTSYT 60  
DB 1 QVQLQSGGGLVQAGGSLRLSCAASGSIIRISIMTWYQAPGKERELVARMSSDGTTSYA 60  
QY 61 DSMKGRFTISRDNKNTVYLMHNNLKPEDTAVYICKALISSYDGSWMDYGGQGTQTVV 120  
DB 61 DSVKGRFTISRDNKNTVYLMHNNLKPEDTAVYICNARRSNYDKRSGWGYGQGTQTVV 120  
QY 121 BPKTPKP 127  
DB 121 AHSSEDP 127  
RESULT 5  
AAE10561  
ID AAE10561 standard; peptide; 129 AA.  
XX  
XX AAE10561;  
XX  
XX 10-DEC-2001 (first entry)  
XX  
XX HGL inhibiting VHH fragment, HGL #4 from llama species.  
XX  
XX Llama antibody; camelid; anorectic; heavy chain variable domain; VHH;  
KW human dietary enzyme inhibitor; medicament; human pancreatic lipase; HPL;  
KW food; human gastric lipase; HGL; cosmetic control; body weight.  
XX  
XX Lama sp.  
XX  
XX Location/Qualifiers  
FH Key 31..35  
FT Region /label= CDR1  
FT /note= "Complementarity determining region 1"  
FT Region 50..64  
FT /label= CDR2  
FT /note= "Complementarity determining region 2"  
FT Region 98..109  
FT /label= CDR3  
FT /note= "Complementarity determining region 3"  
XX  
XX EPI134231-A1.  
XX

```

XX 19-SEP-2001.
XX 20-FEB-2001; 2001EP-00200703.
XX 14-MAR-2000; 2000EP-00200930.
XX (UNIL ) UNILEVER NV.
XX (UNIL ) UNILEVER PLC.
XX Bezemer S, Van De Burg M, De Haard JJW, Tareilus E;
XX WPI; 2001-572718/65.
XX New antibody or its fragments for inhibiting human dietary enzymes,
XX useful for cosmetic control of body weight of human beings, comprises
XX heavy chain variable domain derived from immunoglobulin naturally devoid
XX of light chains.
XX Example 4; Page 13; 37pp; English.
XX The patent discloses antibodies or their fragments comprising a heavy
XX chain variable domain (VHH) derived from an immunoglobulin naturally
XX devoid of light chains specific for inhibiting human dietary enzymes. The
XX antibodies of the invention are useful for the preparation of medicaments
XX or food for inhibiting the activity of one or more human dietary enzymes
XX especially human pancreatic lipase (HPL) or human gastric lipase (HGL)
XX which are useful for the cosmetic control of body weight of human beings.
XX The present peptide sequence is HGL inhibiting VHH fragment, HGL #4 from
XX llama (camelid) species
XX Sequence 129 AA;
Query Match 73.3%; Score 494; DB 4; Length 129;
Best Local Similarity 75.2%; Pred. No. 2.4e-38;
Matches 97; Conservative 10; Mismatches 22; Indels 0; Gaps 0;
OY 1 QVQLQESGGGLVQAGGSLRLSCAASGSTRSISIMTWYRQAPGKERELVARMSSDGTTSYT 60
Db 1 QVQLQESGGGLVQAGGSLRLSCAASGSTRSISIMTWYRQAPGKERELVARMSSDGTTSYT 60
OY 61 DSMKGRFTISRDNKNTVYLHMNNLKPEDTAVYYCKALISSYDGSWNDYGGQGTQVTVSS 120
Db 61 DSVKGRFTIARENKNTVYLQWSSLKPEDTAVYYCKRGLTQYSEHDYWGQGTQVTVSS 120
OY 121 EPKTPKPQP 129
Db 121 EPKTPKPQP 129
RESULT 6
AAE10556
ID AAE10556 standard; peptide; 130 AA.
XX AAE10556;
XX 10-DEC-2001 (first entry)
XX HPL inhibiting VHH fragment, HPL #18 from llama species.
XX Llama antibody; camelid; anorectic; heavy chain variable domain; VHH;
XX human dietary enzyme inhibitor; medicament; human pancreatic lipase; HPL;
XX food; human gastric lipase; HGL; cosmetic control; body weight.
XX Lama sp.
XX Location/Qualifiers
XX Key 31..35
XX Region
XX /label= CDR1
XX /note= "Complementarity determining region 1"
XX Region 50..64
XX /label= CDR2
XX /note= "Complementarity determining region 2"

```

```

FT Region 98..110
FT /label= CDR3
FT /note= "Complementarity determining region 3"
XX EP1134231-A1.
XX 19-SEP-2001.
XX 20-FEB-2001; 2001EP-00200703.
XX 14-MAR-2000; 2000EP-00200930.
XX (UNIL ) UNILEVER NV.
XX (UNIL ) UNILEVER PLC.
XX Bezemer S, Van De Burg M, De Haard JJW, Tareilus E;
XX WPI; 2001-572718/65.
XX New antibody or its fragments for inhibiting human dietary enzymes,
XX useful for cosmetic control of body weight of human beings, comprises
XX heavy chain variable domain derived from immunoglobulin naturally devoid
XX of light chains.
XX Example 2; Page 10; 37pp; English.
XX The patent discloses antibodies or their fragments comprising a heavy
XX chain variable domain (VHH) derived from an immunoglobulin naturally
XX devoid of light chains specific for inhibiting human dietary enzymes. The
XX antibodies of the invention are useful for the preparation of medicaments
XX or food for inhibiting the activity of one or more human dietary enzymes
XX especially human pancreatic lipase (HPL) or human gastric lipase (HGL)
XX which are useful for the cosmetic control of body weight of human beings.
XX The present peptide sequence is HPL inhibiting VHH fragment, HPL #18 from
XX llama (camelid) species
XX Sequence 130 AA;
Query Match 72.9%; Score 491.5; DB 4; Length 130;
Best Local Similarity 73.8%; Pred. No. 4.2e-38;
Matches 96; Conservative 11; Mismatches 22; Indels 1; Gaps 1;
OY 1 QVQLQESGGGLVQAGGSLRLSCAASGSTRSISIMTWYRQAPGKERELVARMSSDGTTSYT 60
Db 1 QVQLQESGGGLVQAGGSLRLSCAASGSTRSISIMTWYRQAPGKERELVARMSSDGTTSYT 60
OY 61 DSMKGRFTISRDNKNTVYLHMNNLKPEDTAVYYCKALISSYDGSWNDYGGQGTQVTVSS 119
Db 61 DSVKGRFTISRDNKNTVYLQWSSLKPEDTAVYYCKALIRKFTSEYNEWYGGQGTQVTVSS 120
OY 120 SEPKTPKPQP 129
Db 121 SEPKTPKPQP 130
RESULT 7
AAE10563
ID AAE10563 standard; peptide; 124 AA.
XX AAE10563;
XX 10-DEC-2001 (first entry)
XX HGL inhibiting VHH fragment, HGL #9 from llama species.
XX Llama antibody; camelid; anorectic; heavy chain variable domain; VHH;
XX human dietary enzyme inhibitor; medicament; human pancreatic lipase; HPL;
XX food; human gastric lipase; HGL; cosmetic control; body weight.
XX Lama sp.
XX Location/Qualifiers
XX Key 31..35
XX Region
XX /label= CDR1
XX /note= "Complementarity determining region 1"
XX Region 50..64
XX /label= CDR2
XX /note= "Complementarity determining region 2"

```



FT /label= CDR1  
 FT /note= "Complementarity determining region 1"  
 FT 50..64  
 FT /label= CDR2  
 FT /note= "Complementarity determining region 2"  
 FT 98..104  
 FT /label= CDR3  
 FT /note= "Complementarity determining region 3"  
 XX  
 PN EP1134231-A1.  
 XX  
 PD 19-SEP-2001.  
 XX  
 PE 20-FEB-2001; 2001EP-00200703.  
 XX  
 PR 14-MAR-2000; 2000EP-00200930.  
 XX  
 PA (UNIL ) UNILEVER NV.  
 PA (UNIL ) UNILEVER PLC.  
 XX  
 PI Bezemer S, Van De Burg M, De Haard JJW, Tareilus E;  
 XX  
 DR WPI; 2001-572718/65.  
 XX

PT New antibody or its fragments for inhibiting human dietary enzymes,  
 PT useful for cosmetic control of body weight of human beings, comprises  
 PT heavy chain variable domain derived from immunoglobulin naturally devoid  
 PT of light chains.

PS Example 4; Page 13; 37pp; English.

CC The patent discloses antibodies or their fragments comprising a heavy  
 CC chain variable domain (VHH) derived from an immunoglobulin naturally  
 CC devoid of light chains specific for inhibiting human dietary enzymes. The  
 CC antibodies of the invention are useful for the preparation of medicaments  
 CC or food for inhibiting the activity of one or more human dietary enzymes,  
 CC especially human pancreatic lipase (HPL) or human gastric lipase (HGL)  
 CC which are useful for the cosmetic control of body weight of human beings.  
 CC The present peptide sequence is HGL inhibiting VHH fragment, HGL #9 from  
 CC llama (camelid) species

XX Sequence 124 AA;

Query Match 72.5%; Score 488.5; DB 4; Length 124;  
 Best Local Similarity 73.6%; Pred. No. 7.6e-38;  
 Matches 95; Conservative 11; Mismatches 18; Indels 5; Gaps 1;

QY 1 QVQLQSGGGLVQAGGSLRLSCAASGSIISIMTWYRQAPGKERELVARMSSDGTTSYT 60  
 DB 1 QVQLQSGGGLVQAGGSLRLSCAASGSIISIMTWYRQAPGKERELVARMSSDGTTSYT 60  
 QY 61 DSMKGRFTISRDNKNTVYLHMNKLKPEDTAVYCKALISSYDGSWNDYGGGQTQVTSS 120  
 DB 61 DSVKGRFTISRDNKNTVYLHMNKLKPEDTAVYCKALISSYDGSWNDYGGGQTQVTSS 115  
 QY 121 EPKTPKPQP 129  
 DB 116 EPKTPKPQP 124

RESULT 8  
 AAE10564  
 ID AAE10564 standard; peptide; 129 AA.  
 XX  
 AC AAE10564;  
 XX  
 DT 10-DEC-2001 (first entry)  
 XX  
 DE HGL inhibiting VHH fragment, HGL #10 from llama species.

XX Llama antibody; camelid; anorectic; heavy chain variable domain; VHH;  
 KW human dietary enzyme inhibitor; medicament; human pancreatic lipase; HPL;  
 KW food; human gastric lipase; HGL; cosmetic control; body weight.

XX Lama sp.  
 OS  
 XX key  
 FH Region  
 FT 31..35  
 FT /label= CDR1  
 FT /note= "Complementarity determining region 1"  
 FT 50..64  
 FT /label= CDR2  
 FT /note= "Complementarity determining region 2"  
 FT 98..109  
 FT /label= CDR3  
 FT /note= "Complementarity determining region 3"

PN EP1134231-A1.

XX

PD 19-SEP-2001.

XX

PE 20-FEB-2001; 2001EP-00200703.

XX

PR 14-MAR-2000; 2000EP-00200930.

XX

PA (UNIL ) UNILEVER NV.

PA (UNIL ) UNILEVER PLC.

XX

PI Bezemer S, Van De Burg M, De Haard JJW, Tareilus E;

XX

DR WPI; 2001-572718/65.

XX

PT New antibody or its fragments for inhibiting human dietary enzymes,

PT useful for cosmetic control of body weight of human beings, comprises

PT heavy chain variable domain derived from immunoglobulin naturally devoid

PT of light chains.

XX

PS Example 4; Page 13; 37pp; English.

XX

CC The patent discloses antibodies or their fragments comprising a heavy

CC chain variable domain (VHH) derived from an immunoglobulin naturally

CC devoid of light chains specific for inhibiting human dietary enzymes. The

CC antibodies of the invention are useful for the preparation of medicaments

CC or food for inhibiting the activity of one or more human dietary enzymes

CC especially human pancreatic lipase (HPL) or human gastric lipase (HGL)

CC which are useful for the cosmetic control of body weight of human beings.

CC The present peptide sequence is HGL inhibiting VHH fragment, HGL #10 from

CC llama (camelid) species

XX Sequence 129 AA;

Query Match 72.4%; Score 488; DB 4; Length 129;

Best Local Similarity 74.4%; Pred. No. 8.8e-38;

Matches 96; Conservative 11; Mismatches 22; Indels 0; Gaps 0;

QY 1 QVQLQSGGGLVQAGGSLRLSCAASGSIISIMTWYRQAPGKERELVARMSSDGTTSYT 60

DB 1 QVQLQSGGGLVQAGGSLRLSCAASGSIISIMTWYRQAPGKERELVARMSSDGTTSYT 60

QY 61 DSMKGRFTISRDNKNTVYLHMNKLKPEDTAVYCKALISSYDGSWNDYGGGQTQVTSS 120

DB 61 NSVKGRTISRDNKNTVYLQMNSLKPEDTAVYCKGRGLTQYSEHDIWGGQTQVTSS 120

QY 121 EPKTPKPQP 129

DB 121 EPKTPKPQP 129

RESULT 9

AAE10557

ID AAE10557 standard; peptide; 130 AA.

XX

AC AAE10557;

XX

DT 10-DEC-2001 (first entry)

XX

```
DE HPL inhibiting VHH fragment, HPL #19 from llama species.
XX
XX Llama antibody; camelid; anorectic; heavy chain variable domain; VHH;
KW human dietary enzyme inhibitor; medicament; human pancreatic lipase; HPL;
KW food; human gastric lipase; HGL; cosmetic control; body weight.
XX
OS Lama sp.
XX
XX Location/Qualifiers
FH Key 31..35
FT Region /label= CDR1
FT /note= "Complementarity determining region 1"
FT Region 50..64
FT /label= CDR2
FT /note= "Complementarity determining region 2"
FT Region 98..110
FT /label= CDR3
FT /note= "Complementarity determining region 3"
XX
XX EP1134231-A1.
XX
XX 19-SEP-2001.
XX
XX 20-FEB-2001; 2001EP-00200703.
XX
XX 14-MAR-2000; 2000EP-00200930.
XX
XX (UNIL ) UNILEVER NV.
PA (UNIL ) UNILEVER PLC.
XX
XX Bezemer S, Van De Burg M, De Haard JJW, Tareilus E;
XX WPI; 2001-572718/65.
XX
XX New antibody or its fragments for inhibiting human dietary enzymes,
PT useful for cosmetic control of body weight of human beings, comprises
PT heavy chain variable domain derived from immunoglobulin naturally devoid
PT of light chains..
XX
XX Example 2; Page 10; 37pp; English.
XX
XX The patent discloses antibodies or their fragments comprising a heavy
CC chain variable domain (VHH) derived from an immunoglobulin naturally
CC devoid of light chains specific for inhibiting human dietary enzymes. The
CC antibodies of the invention are useful for the preparation of medicaments
CC or food for inhibiting the activity of one or more human dietary enzymes
CC especially human pancreatic lipase (HPL) or human gastric lipase (HGL)
CC which are useful for the cosmetic control of body weight of human beings.
CC The present peptide sequence is HPL inhibiting VHH fragment, HPL #19 from
CC llama (camelid) species
XX
XX Sequence 130 AA;
XX
XX Query Match 72.3%; Score 487.5; DB 4; Length 130;
XX Best Local Similarity 71.5%; Pred. No. 9.9e-38;
XX Matches 93; Conservative 14; Mismatches 22; Indels 1; Gaps 1;
XX
XX 1 QVQLQESGGGLVQAGGSLRLSCAASGSIISIMTWYRQAPGKERELVARMSSDGTTSYT 60
XX 1 QVQLQDSGGGLVQAGGSLRLSCAASGPIGVLMGYRQAPGKQREWVASITATGPNYIT 60
XX
XX 61 DSMKGRFTISRDNAKNTVYLHMNLIKPEDTAVVYCKALIS-SYDGSWNDYGGOGTQTVTS 119
XX 61 DSVKGRFTISRDNKNTVYLMNLIKPEDTAVVYCNQAVRFRSSDYTYNWGGGTQTVTS 120
XX
XX 120 SEPTPKPQP 129
XX 121 SEPTPKPQP 130
XX
XX RESULT 10.
XX AAE10566
XX ID AAE10566 standard; peptide; 124 AA.
```

```
XX
XX AAE10566;
XX
XX 10-DEC-2001 (first entry)
XX
XX HGL inhibiting VHH fragment, HGL #15 from llama species.
XX
XX Llama antibody; camelid; anorectic; heavy chain variable domain; VHH;
KW human dietary enzyme inhibitor; medicament; human pancreatic lipase; HPL;
KW food; human gastric lipase; HGL; cosmetic control; body weight.
XX
XX Lama sp.
XX
XX Location/Qualifiers
FH Key 30..35
FT Region /label= CDR1
FT /note= "Complementarity determining region 1"
FT Region 50..64
FT /label= CDR2
FT /note= "Complementarity determining region 2"
FT Region 98..104
FT /label= CDR3
FT /note= "Complementarity determining region 3"
XX
XX EP1134231-A1.
XX
XX 19-SEP-2001.
XX
XX 20-FEB-2001; 2001EP-00200703.
XX
XX 14-MAR-2000; 2000EP-00200930.
XX
XX (UNIL ) UNILEVER NV.
PA (UNIL ) UNILEVER PLC.
XX
XX Bezemer S, Van De Burg M, De Haard JJW, Tareilus E;
XX WPI; 2001-572718/65.
XX
XX New antibody or its fragments for inhibiting human dietary enzymes,
PT useful for cosmetic control of body weight of human beings, comprises
PT heavy chain variable domain derived from immunoglobulin naturally devoid
PT of light chains.
XX
XX Example 4; Page 14; 37pp; English.
XX
XX The patent discloses antibodies or their fragments comprising a heavy
CC chain variable domain (VHH) derived from an immunoglobulin naturally
CC devoid of light chains specific for inhibiting human dietary enzymes. The
CC antibodies of the invention are useful for the preparation of medicaments
CC or food for inhibiting the activity of one or more human dietary enzymes
CC especially human pancreatic lipase (HPL) or human gastric lipase (HGL)
CC which are useful for the cosmetic control of body weight of human beings.
CC The present peptide sequence is HGL inhibiting VHH fragment, HGL #15 from
CC llama (camelid) species
XX
XX Sequence 124 AA;
XX
XX Query Match 71.7%; Score 483.5; DB 4; Length 124;
XX Best Local Similarity 72.9%; Pred. No. 2.2e-37;
XX Matches 94; Conservative 11; Mismatches 19; Indels 5; Gaps 1;
XX
XX 1 QVQLQESGGGLVQAGGSLRLSCAASGSIISIMTWYRQAPGKERELVARMSSDGTTSYT 60
XX 1 QVQLQESGGGLVQAGGSLRLSCAASGSIISIMTWYRQAPGKERELVARMSSDGTTSYT 60
XX
XX 61 DSMKGRFTISRDNAKNTVYLHMNLIKPEDTAVVYCKALIS-SYDGSWNDYGGOGTQTVTS 120
XX 61 DSVKGRFTISRDNKNTVYLMNLIKPEDTAVVYCNQAVRFRSSDYTYNWGGGTQTVTS 115
XX
XX 121 EPKTPKQP 129
XX 116 EPKTPKQP 124
```

Db	61	DSVGRFTTSIDNTLVYLMQNDLKPEDTAVYCNADVPRYTSRYLEWVGQGLTVTS	120
QY	120	SEPTKPEQP	129
Db	121	SEPTKPEQP	130
RESULT 12			
AAE05283	ID AAE05283 standard; protein; 152 AA.		
XX	AC		
XX	AAE05283;		
DT	18-SEP-2001	(first entry)	
DE	Anti-potato SBEII (Clone46)	VH region attached with myc and his6 tag.	
XX	Potato; heavy chain immunoglobulin; pathogen resistance;		
KW	metabolism modulator; passive immunisation; heavy chain variable domain;		
XW	VH; anti-potato SBEII; starch branching enzyme; SBE A.		
XX	Solanum tuberosum.		
OS	Unidentified.		
OS	Chimeric.		
PN	EP1118669-A2.		
XX	25-JUL-2001.		
PD	08-DEC-2000; 2000EP-00310997.		
PF	17-DEC-1999; 99EP-00310188.		
PR	(UNIL ) UNILEVER PLC.		
XX	(UNIL ) UNILEVER NV.		
PA	Frenken LGJ, Van Der Logt CPE, Jobling SA, Teh Y;		
PI	WPI: 2001-427157/46.		
XX	N-PSGB; AAD10054.		
DR	Modifying a plant to produce an antibody useful for increasing pathogen		
PT	resistance or to modulate metabolism comprises introducing a DNA sequence		
PT	encoding a heavy chain immunoglobulin linked to a peptide that targets a		
PT	cellular compartment.		
PS	Example 1; Fig 12; 81pp; English.		
XX	The present invention relates to a method for modifying a plant to		
CC	produce an antibody or an active fragment or derivative, or a protein		
CC	functional equivalent, in a cellular compartment comprises introducing a		
CC	DNA sequence encoding a heavy chain immunoglobulin, where the DNA is		
CC	linked to promoters and provided with an additional sequence encoding a		
CC	peptide capable of targeting heavy chain immunoglobulin to a cellular		
CC	compartment. The method is used for producing a heavy chain		
CC	immunoglobulin or an active fragment or derivative, or a protein that is		
CC	functionally equivalent for increasing the pathogen resistance in a plant		
CC	or to modulate metabolism in a plant. Under some circumstances it may be		
CC	desirable to retain the antibody product with the plant rather than		
CC	extracting and isolating the product. In particular, edible selected		
CC	antigens may be used in a method of passively immunising an animal,		
CC	preferably human, against the antigen, e.g., pathogenic organisms. The		
CC	present sequence is an anti-potato SBEII antibody (denoted Clone46) heavy		
CC	chain variable domain (VH) attached to peptide linkers, myc and his6 tag		
CC	The potato SBEII is a starch branching enzyme also designated SBE A		
XX	Sequence 152 AA;		
SQ			

```

Db 1 QVQLQESGGGLVQAGGSLRLSCVASGNTFSIAAWYRQAPGKQREVVASINSIGSTNYA 60
QY 61 DSMKGRFTISRDNKNTVYLNHNKLPEDTAVYYCKA---LISSYDGSWNDYGGQGTQVT 117
Db 61 DSVKGRFTISRDNKNTVYLNHNKLPEDTAVYYCAAGNLLVK-----RPYWGQGTTLVT 114
QY 118 VSSEPKTKPQP 129
Db 115 VSSEPKTKPQP 126

RESULT 13
AAE05288
ID AAE05288 standard; protein; 149 AA.
XX
AC AAE05288;
XX
DT 18-SEP-2001 (first entry)
XX
DE Anti-potato SBEII (Clone68) VH region attached with myc and his6 tag.
XX
KW Potato; heavy chain immunoglobulin; pathogen resistance;
KW metabolism modulator; passive immunisation; heavy chain variable domain;
KW anti-potato SBEII; starch branching enzyme; SBE A.
XX
OS Solanum tuberosum.
OS Unidentified.
OS Chimeric.
XX
PN EP1118669-A2.
XX
PD 25-JUL-2001.
XX
PF 08-DEC-2000; 2000EP-00310997.
XX
PR 17-DEC-1999; 99EP-00310188.
XX
PA (UNIL ) UNILEVER PLC.
PA (UNIL ) UNILEVER NV.
XX
PI Frenken LGJ, Van Der Logt CPE, Jobling SA, Teh Y;
XX
WPI; 2001-427157/46.
DR N-PSDB; AAD10059.
XX
PT Modifying a plant to produce an antibody useful for increasing pathogen
PT resistance or to modulate metabolism comprises introducing a DNA sequence
PT encoding a heavy chain immunoglobulin linked to a peptide that targets a
PT cellular compartment.
XX
PS Example 13; Fig 28; 81pp; English.
XX
CC The present invention relates to a method for modifying a plant to
CC produce an antibody or an active fragment or derivative, or a protein
CC functional equivalent, in a cellular compartment comprising introducing a
CC DNA sequence encoding a heavy chain immunoglobulin, where the DNA is
CC linked to promoters and provided with an additional sequence encoding a
CC peptide capable of targeting heavy chain immunoglobulin to a cellular
CC compartment. The method is used for producing a heavy chain
CC immunoglobulin or an active fragment or derivative, or a protein that is
CC functionally equivalent for increasing the pathogen resistance in a plant
CC or to modulate metabolism in a plant. Under some circumstances it may be
CC desirable to retain the antibody product with the plant rather than
CC extracting and isolating the product. In particular, edible selected
CC antigens may be used in a method of passively immunising an animal,
CC preferably human, against the antigen, e.g., pathogenic organisms. The
CC present sequence is an anti-potato SBEII antibody (denoted Clone68) heavy
CC chain variable domain (VH) attached to peptide linkers, myc and his6 tag.
CC The potato SBEII is a starch branching enzyme also designated SBE A
XX
SQ Sequence 149 AA;

```

```

Query Match 70.3%; Score 474; DB 4; Length 149;
Best Local Similarity 76.0%; Pred. No. 2.2e-36;
Matches 98; Conservative 8; Mismatches 15; Indels 8; Gaps 2;

QY 1 QVQLQESGGGLVQAGGSLRLSCAASGSISSIMTWYRQAPGKERELVARMSDQTTST 60
Db 3 QVQLQESGGGLVQAGGSLRLSCAASGSISSIMTWYRQAPGKERELVALISAGRTWYA 62
QY 61 DSMKGRFTISRDNKNTVYLNHNKLPEDTAVYYCKALISSYDGSWNDYGGQGTQVT 120
Db 63 DSVKGRFTISRDNKNTVYLNHNKLPEDTAVYYCTA-----GGS---YWGQGTQVT 114
QY 121 EPKTPKQP 129
Db 115 EPKTPKQP 123

RESULT 14
AAE10567
ID AAE10567 standard; peptide; 130 AA.
XX
AC AAE10567;
XX
DT 10-DEC-2001 (first entry)
XX
DE HGL inhibiting VHH fragment, HGL #16 from llama species.
XX
KW Llama antibody; camelid; anorectic; heavy chain variable domain; VHH;
KW human dietary enzyme inhibitor; medicament; human pancreatic lipase; HPL;
KW food; human gastric lipase; HGL; cosmetic control; body weight.
XX
OS Lama sp.
XX
FH Key
FT Region 31.35
FT /label= CDR1
FT /note= "Complementarity determining region 1"
FT Region 50.64
FT /label= CDR2
FT /note= "Complementarity determining region 2"
FT Region 98.110
FT /label= CDR3
FT /note= "Complementarity determining region 3"
XX
EP1134231-A1.
XX
PN
XX
PD 19-SEP-2001.
XX
PF 20-FEB-2001; 2001EP-00200703.
XX
PR 14-MAR-2000; 2000EP-00200930.
XX
PA (UNIL ) UNILEVER NV.
PA (UNIL ) UNILEVER PLC.
XX
PI Bezemer S, Van De Burg M, De Haard JJW, Tareilus E;
XX
WPI; 2001-572718/65.
XX
PT New antibody or its fragments for inhibiting human dietary enzymes,
PT useful for cosmetic control of body weight of human beings, comprises
PT heavy chain variable domain derived from immunoglobulin naturally devoid
PT of light chains.
XX
PS Example 4; Page 14; 37pp; English.
XX
CC The patent discloses antibodies or their fragments comprising a heavy
CC chain variable domain (VHH) derived from an immunoglobulin naturally
CC devoid of light chains specific for inhibiting human dietary enzymes. The
CC antibodies of the invention are useful for the preparation of medicaments
CC or food for inhibiting the activity of one or more human dietary enzymes
CC especially human pancreatic lipase (HPL) or human gastric lipase (HGL)
CC which are useful for the cosmetic control of body weight of human beings.

```

CC The present peptide sequence is HGL inhibiting VHH fragment, HGL #16 from  
 CC llama (camelid) species  
 XX  
 SQ Sequence 130 AA;

Query Match 69.7%; Score 469.5; DB 4; Length 130;  
 Best Local Similarity 71.9%; Pred. No. 4.8e-36;  
 Matches 97; Conservative 9; Mismatches 18; Indels 11; Gaps 2;

QY 1 QVQLQESGGGLVQAGGSLRLSCAASGSIIRSIIMTWYRQAPGKERELVARMSSDGTTSYT 60  
 Db 1 QVQLQESGGGLVQAGGSLRLSCAASGSDPRYAMAWYRQAPGKRLVATITVTRTNYA 60

QY 61 DSMKGRFTISRDNKNTVYLMNNLKPEDTAVYICKALIS-----SYDGSWNDYGGQGT 114  
 Db 61 DSVKGRFTISRDNKNTVYLMNNLKPEDTAVYICAAESLEITPTSY-----DYWGQGT 115

QY 115 QVTVSSEPTKPPQ 129  
 Db 116 QVTVSSEPTKPPQ 130

## RESULT 15

AAE10554  
 ID AAE10554 standard; peptide; 130 AA.  
 XX  
 AC AAE10554;  
 XX  
 DT 10-DEC-2001 (first entry)  
 XX  
 DE HPL inhibiting VHH fragment, HPL #14 from llama species.  
 XX  
 KW Llama antibody; camelid; anorectic; heavy chain variable domain; VHH;  
 KW human dietary enzyme inhibitor; medicament; human pancreatic lipase; HPL;  
 KW food; human gastric lipase; HGL; cosmetic control; body weight.  
 XX  
 OS Lama sp.

Key	Location/Qualifiers
FT Region	31..35
FT	/label= CDR1
FT	/note= "Complementarity determining region 1"
FT Region	50..64
FT	/label= CDR2
FT	/note= "Complementarity determining region 2"
FT Region	98..110
FT	/label= CDR3
FT	/note= "Complementarity determining region 3"
XX	
PN	EP1134231-A1.
XX	
PD	19-SEP-2001.
XX	
PF	20-FEB-2001; 2001EP-00200703.
XX	
PR	14-MAR-2000; 2000EP-00200930.
XX	
PA	(UNIL ) UNILEVER NV.
PA	(UNIL ) UNILEVER PLC.
XX	
PI	Bezemer S, Van De Burg M, De Haard JMW, Tareilus E;
XX	
DR	WPI; 2001-572718/65.
XX	
PT	New antibody or its fragments for inhibiting human dietary enzymes,
FT	useful for cosmetic control of body weight of human beings, comprises
PT	heavy chain variable domain derived from immunoglobulin naturally devoid
PT	of light chains.
XX	

Example 2; Page 9; 37pp; English.

PS The patent discloses antibodies or their fragments comprising a heavy  
 CC chain variable domain (VHH) derived from an immunoglobulin naturally

CC devoid of light chains specific for inhibiting human dietary enzymes. The  
 CC antibodies of the invention are useful for the preparation of medicaments  
 CC or food for inhibiting the activity of one or more human dietary enzymes  
 CC especially human pancreatic lipase (HPL) or human gastric lipase (HGL)  
 CC which are useful for the cosmetic control of body weight of human beings.  
 CC The present peptide sequence is HPL inhibiting VHH fragment, HPL #14 from  
 CC llama (camelid) species  
 XX

SQ Sequence 130 AA;

Query Match 69.4%; Score 467.5; DB 4; Length 130;  
 Best Local Similarity 72.3%; Pred. No. 7.3e-36;  
 Matches 94; Conservative 11; Mismatches 24; Indels 1; Gaps 1;

QY 1 QVQLQESGGGLVQAGGSLRLSCAASGSIIRSIIMTWYRQAPGKERELVARMSSDGTTSYT 60  
 Db 1 QVQLQESGGGLVQAGGSLRLSCAASGSIIRSIIMTWYRQAPGKERELVARMSSDGTTSYT 60

QY 61 DSMKGRFTISRDNKNTVYLMNNLKPEDTAVYICKALISYDGS-WNDYGGQGTQVTVS 119  
 Db 61 DSVKGRFTISRDNKNTVYLMNNLKPEDTAVYICAAESLEITPTSY-----DYWGQGT 120

QY 120 SEPTKPPQ 129  
 Db 121 SEPTKPPQ 130

Search completed: October 5, 2004, 08:00:13  
 Job time : 98.1435 secs

This Page Blank (USPIC)

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2004, 07:36:16 ; Search time 21.0648 Seconds  
(without alignments)  
593.639 Million cell updates/sec

Title: US-09-805-290A-26

Perfect score: 692

Sequence: 1 QVQLDSGGGLVQAGGSLRL.....WGQGTLVTVFLEPKTPKQP 130

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR 78:\*\*

1: pir1:\*\*

2: pir2:\*\*

3: pir3:\*\*

4: pir4:\*\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	362	52.3	119	2 S31107	Ig heavy chain - h
2	360	52.0	140	2 S31588	Ig heavy chain V r
3	359	51.9	119	2 S31108	Ig heavy chain - h
4	358.5	51.8	120	2 S48798	Ig heavy chain V r
5	358	51.7	119	2 D38005	Ig heavy chain V r
6	355	51.3	127	2 S38489	Ig heavy chain - h
7	355	51.3	138	2 S31666	Ig heavy chain V r
8	354	51.2	160	2 S05271	Ig heavy chain pre
9	352	50.9	121	2 I55673	Ig heavy chain - h
10	352	50.9	123	2 S31114	Ig heavy chain - h
11	351	50.7	119	2 C36005	Ig heavy chain V r
12	349	50.4	122	2 E27888	Ig heavy chain V r
13	347	50.1	121	2 S26798	Ig heavy chain V r
14	347	50.1	123	2 S26794	Ig heavy chain V r
15	346.5	50.1	116	2 S31110	Ig heavy chain - h
16	346	50.0	120	2 S36278	Ig heavy chain V r
17	345	49.9	120	2 S55536	Ig heavy chain V r
18	345	49.9	125	2 S30531	Ig heavy chain V r
19	344.5	49.8	120	2 S44111	Ig heavy chain V r
20	344	49.7	134	2 S31699	Ig heavy chain V r
21	343	49.6	121	2 S31113	Ig heavy chain V r
22	343	49.6	140	2 S31686	Ig heavy chain - h
23	342.5	49.5	124	2 S20782	Ig heavy chain V r
24	342.5	49.5	147	2 I37780	Ig variable region
25	339.5	49.1	128	2 S26786	Ig heavy chain V r
26	338	49.0	138	2 S09258	Ig heavy chain V r
27	338.5	48.9	140	2 S70442	Ig heavy chain pre
28	338	48.8	119	2 F36005	Ig heavy chain V r
29	338	48.8	120	2 S55537	Ig heavy chain V r

30 338 48.8 127 2 S19878 Ig heavy chain V r  
31 337.5 48.8 114 2 S46391 Ig heavy chain V r  
32 337 48.7 139 2 S31678 Ig heavy chain V r  
33 336.5 48.6 112 2 P31647 Ig heavy chain V r  
34 336.5 48.6 128 2 S48797 Ig heavy chain pre  
35 336.5 48.6 151 2 A60943 Ig heavy chain V r  
36 336 48.6 135 2 S31598 Ig heavy chain V r  
37 335 48.4 115 1 A3H8U0 Ig heavy chain V r  
38 335 48.4 121 2 S31104 Ig heavy chain (su  
39 335 48.4 138 2 A30561 Ig heavy chain pre  
40 335 48.4 140 2 S22657 Ig heavy chain pre  
41 334.5 48.3 114 2 S31120 Ig heavy chain - h  
42 334.5 48.3 119 2 E27888 Ig heavy chain V r  
43 334.5 48.3 119 2 D27889 Ig heavy chain V r  
44 334.5 48.3 128 2 S26790 Ig heavy chain V r  
45 334 48.3 117 2 S31109 Ig heavy chain - h

#### ALIGNMENTS

##### RESULT 1

S31107

Ig heavy chain - human

C:Species: Homo sapiens (man)

C>Date: 02-Dec-1993 #sequence\_revision 26-May-1995 #text\_change 17-Mar-1999

C:Accession: S31107

R:Raaphorst, F.M.; Timmers, E.; Kenter, M.J.H.; van Tol, M.J.D.; Vossen, J.M.; Schuurman,

Eur. J. Immunol. 22, 247-251, 1992

A:Title: Restricted utilization of germ-line V(H)3 genes and short diverse third comple

A:Reference number: S31104; MUID:92111633; PMID:1730252

A:Accession: S31107

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: mRNA

A:Residues: 1-119 <RA>

A:Cross-references: EMBL:X62955

A>Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1991

C:Superfamily: immunoglobulin V region; immunoglobulin homology

C:Keywords: heterotetramer; immunoglobulin

F:15-98/Domain: immunoglobulin homology <IMM>

Query Match 52.3%; Score 362; DB 2; Length 119;

Best Local Similarity 64.2%; Pred. No. 5.7e-25;

Matches 77; Conservative 9; Mismatches 30; Indels 4; Gaps 2;

QY 1 QVQLDSGGGLVQAGGSLRLSCAASGSIQDHTMGWYRQTPGKQDVVATIQ-SGGSTNY 59

Db 1 EVQLVESGGGLVQPGGSLRLSCAASGTFPSYAMSWRQAPGKLEWVSAISGSGSTYY 60

QY 60 ADSVKGRFTISRDNLTNTVYIQMNDLKPEDTGVYWNADVRPVRTSYLBIWGGTLTV 119

Db 61 ADSVKGRFTISRDNLSKNTLVQMNSLRADTAVYCAKDP---GASYFDYWGQGTLLTV 117

##### RESULT 2

S31588

Ig heavy chain V region - human (fragment)

C:Species: Homo sapiens (man)

C>Date: 22-Nov-1993 #sequence\_revision 10-Nov-1995 #text\_change 23-Jul-1999

C:Accession: S31588

R:Cuisinier, A.M.; Gauthier, L.; Boublil, L.; Fougereau, M.; Tonnelie, C.

submitted to the EMBL Data Library, June 1992

A:Description: Mechanisms that generate human immunoglobulin diversity operate from the {

A:Reference number: S31585

A:Accession: S31588

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-140 <CUI>

A:Cross-references: EMBL:Z14200; NID:g30957; PIDN:CAA78569.1; PID:g30958

C:Superfamily: immunoglobulin V region; immunoglobulin homology

C:Keywords: heterotetramer; immunoglobulin

F:34-117/Domain: immunoglobulin homology <IMM>





```

Db      61 ADSVKGRFTISRDNKNTLYLQNSLRADTAVYIC-AKEGFPASDYDSSGYISFDYV 119
      112 GQGLTVTV 119
      120 GQGLTVTV 127

RESULT 7
S31666
Ig heavy chain V region - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 22-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 23-Jul-1999
C:Accession: S31666
R:Cuissinier, A.M.; Gauthier, L.; Boubli, L.; Fougereau, M.; Tonnelie, C.
submitted to the EMBL Data Library, June 1992
A:Description: Mechanisms that generate human immunoglobulin diversity operate from the
A:Reference number: S31585
A:Accession: S31666
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-138 <CU>
A:Cross-references: EMBL:Z14202; NID:g30963; PIDN:CAA78571.1; PID:g30964
C:Superfamily: immunoglobulin V region; immunoglobulin homology
C:Keywords: heterotrimer; immunoglobulin
F:34-117/Domain: immunoglobulin homology <IMM>

Query Match      51.3%; Score 355; DB 2; Length 138;
Best Local Similarity 58.6%; Pred. No. 2,8e-24;
Matches 75; Conservative 11; Mismatches 22; Indels 20; Gaps 3;

QY      1 QVQLDGGGLVQAGSLRSLCAASGSGIDHTMGWYRTPGKQRDVVATIQ-SGGSTNY 59
      20 EVQLLESGGGLVQPGGSLRSLCAASGFTFTSYAMSVWRQAPGKLEWVSAISGGSGTYY 79
      60 ADSVKGRFTISRDNKNTLYLQNSLRADTAVYIC-AKEGFPASDYDSSGYISFDYV 119
      80 ADSVKGRFTISRDNKNTLYLQNSLRADTAVYIC-AKEGFPASDYDSSGYISFDYV 127

RESULT 8
S05271
Ig heavy chain precursor - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 16-Aug-1996
C:Accession: S05271; S04602
R:Kishimoto, T.
submitted to the EMBL Data Library, March 1989
A:Reference number: S05270
A:Accession: S05271
A:Molecule type: mRNA
A:Residues: 1-160 <KIS>
A:Cross-references: EMBL:X14584
R:Kishimoto, T.; Okajima, H.; Okumoto, T.; Taniguchi, M.
Nucleic Acids Res. 17, 4385, 1989
A:Title: Nucleotide sequences of the cDNAs encoding the V-regions of H- and L-chains of
A:Reference number: S04601; MUID:89296497; PMID:2500644
A:Accession: S04602
A:Molecule type: mRNA
A:Residues: 1-144 <KIS2>
A:Cross-references: EMBL:X14584
C:Superfamily: immunoglobulin V region; immunoglobulin homology
C:Keywords: heterotrimer; immunoglobulin
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-160/Product: Ig heavy chain (fragment) #status predicted <MAT>
F:34-117/Domain: immunoglobulin homology <IMM>

Query Match      51.2%; Score 354; DB 2; Length 160;
Best Local Similarity 62.6%; Pred. No. 4e-24;

```

```

Matches 77; Conservative 12; Mismatches 30; Indels 4; Gaps 3;

QY      1 QVQLDGGGLVQAGSLRSLCAASGSGIDHTMGWYRTPGKQRDVVATIQ-SGGSTNY 59
      20 EVQLLESGGGLVQPGGSLRSLCAASGFTFTSYAMSVWRQAPGKLEWVSAISGGSGTYY 79
      60 ADSVKGRFTISRDNKNTLYLQNSLRADTAVYIC-AKEGFPASDYDSSGYISFDYV 119
      80 ADSVKGRFTISRDNKNTLYLQNSLRADTAVYIC-AKEGFPASDYDSSGYISFDYV 127

RESULT 9
S55673
Ig heavy chain - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 21-Jan-2000
C:Accession: I55673
R:Knight, G.B.; Agnello, V.; Bonagura, V.; Barnes, J.L.; Panka, D.J.; Zhang, Q.X.
J. Exp. Med. 178, 1903-1911, 1993
A:Title: Human rheumatoid factor cross-idiotypes. IV. Studies on WA X1d-positive IgM with
tinct from the 17.109 and G6 X1ds.
A:Reference number: I55673; MUID:94065558; PMID:8245772
A:Accession: I55673
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-121 <RES>
A:Cross-references: GB:M87268; NID:g186197; PIDN:AAC37536.1; PID:g186198
C:Genetics:
A:Gene: GDB:IGHM
A:Map position: 14q32.33-14q32.33
A:Cross-references: GDB:I20086; OMIM:147020
C:Superfamily: immunoglobulin V region; immunoglobulin homology
F:15-98/Domain: immunoglobulin homology <IMM>

Query Match      50.9%; Score 352; DB 2; Length 121;
Best Local Similarity 63.3%; Pred. No. 4.4e-24;
Matches 76; Conservative 10; Mismatches 32; Indels 2; Gaps 2;

QY      1 QVQLDGGGLVQAGSLRSLCAASGSGIDHTMGWYRTPGKQRDVVATIQ-SGGSTNY 59
      20 EVQLLESGGGLVQPGGSLRSLCAASGFTFTSYAMSVWRQAPGKLEWVSAISGGSGTYY 60
      60 ADSVKGRFTISRDNKNTLYLQNSLRADTAVYIC-AKEGFPASDYDSSGYISFDYV 119
      61 ADSVKGRFTISRDNKNTLYLQNSLRADTAVYIC-AKEGFPASDYDSSGYISFDYV 119

RESULT 10
S31114
Ig heavy chain - human
C:Species: Homo sapiens (man)
C:Date: 02-Dec-1993 #sequence_revision 26-May-1995 #text_change 17-Mar-1999
C:Accession: S31114
R:Rasaphorst, F.M.; Timmers, E.; Kenter, M.J.H.; van Tol, M.J.D.; Vossen, J.M.; Schuurman,
Eur. J. Immunol. 22, 247-251, 1992
A:Title: Restricted utilization of germ-line V(H)3 genes and short diverse third comple
A:Reference number: S31104; MUID:92111633; PMID:1730252
A:Accession: S31114
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: mRNA
A:Residues: 1-123 <RAA>
A:Cross-references: EMBL:X62963
A:Note: The nucleotide sequence was submitted to the EMBL Data Library, October 1991
C:Superfamily: immunoglobulin V region; immunoglobulin homology
C:Keywords: heterotrimer; immunoglobulin
F:15-98/Domain: immunoglobulin homology <IMM>

Query Match      50.9%; Score 352; DB 2; Length 123;
Best Local Similarity 60.6%; Pred. No. 4.5e-24;

```





This Page Blank (uspto)

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2004, 07:29:35 : Search time 11.6358 Seconds  
(without alignments)  
581.749 Million cell updates/sec

Title: US-09-805-290A-26

Perfect score: 692

Sequence: 1 QVQLDSGGGLVQAGGSLRL.....WGQGITLVTFLEPKTPKQP 130

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	335	48.4	115	1	HV3F HUMAN
2	321	46.4	116	1	HV05 CARAU
3	318.5	46.0	122	1	HV3C HUMAN
4	317.5	45.9	119	1	HV3U HUMAN
5	316.5	45.7	120	1	HV3U HUMAN
6	314	45.4	119	1	HV3I HUMAN
7	313.5	45.3	114	1	HV3B HUMAN
8	312.5	45.2	117	1	HV3C HUMAN
9	312.5	45.2	117	1	HV3C HUMAN
10	311.5	45.0	126	1	HV3K HUMAN
11	306.5	44.3	116	1	HV3T HUMAN
12	305	44.1	115	1	HV33 MOUSE
13	304	43.9	120	1	HV3E HUMAN
14	303.5	43.9	136	1	HV16 MOUSE
15	303	43.8	115	1	HV32 MOUSE
16	302	43.6	117	1	HV02 CANFA
17	301.5	43.6	122	1	HV3A HUMAN
18	301	43.5	113	1	HV27 MOUSE
19	300	43.4	97	1	HV56 MOUSE
20	300	43.4	119	1	HV31 HUMAN
21	299.5	43.3	119	1	HV3P HUMAN
22	299	43.2	115	1	HV3D HUMAN
23	297	42.9	113	1	HV30 MOUSE
24	296.5	42.8	117	1	HV3O HUMAN
25	296	42.8	113	1	HV31 MOUSE
26	296	42.8	121	1	HV3J HUMAN
27	295.5	42.7	111	1	HV35 MOUSE
28	295.5	42.7	117	1	HV55 MOUSE
29	295	42.6	113	1	HV28 MOUSE
30	294.5	42.6	117	1	HV53 MOUSE
31	293.5	42.4	117	1	HV54 MOUSE
32	291.5	42.1	98	1	HV57 MOUSE
33	291.5	42.1	114	1	HV01 CANFA

34	291	42.1	113	1	HV29 MOUSE	P01798 mus musculu
35	289	41.8	123	1	HV24 MOUSE	P01793 mus musculu
36	287.5	41.5	122	1	HV20 MOUSE	P01789 mus musculu
37	286.5	41.4	119	1	HV37 MOUSE	P01807 mus musculu
38	285	41.2	123	1	HV23 MOUSE	P01792 mus musculu
39	283	40.9	123	1	HV25 MOUSE	P01794 mus musculu
40	283	40.9	142	1	HV01 RAT	P01805 xattu norv
41	283	40.9	144	1	HV26 MOUSE	P01795 mus musculu
42	282.5	40.8	117	1	HV58 MOUSE	P18529 mus musculu
43	282	40.8	113	1	HV34 MOUSE	P01803 mus musculu
44	282	40.8	123	1	HV22 MOUSE	P01791 mus musculu
45	281.5	40.7	119	1	HV40 MOUSE	P01810 mus musculu

ALIGNMENTS

RESULT 1

HV3F HUMAN

ID HV3F HUMAN STANDARD; PRT; 115 AA.

AC P01767;

DT 21-JUL-1986 (Rel. 01, Created)

DT 21-JUL-1986 (Rel. 01, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Ig heavy chain V-III region BUT.

OS Homo sapiens (Human)

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBT\_TaxID=9606;

RN [1]

RP SEQUENCE.

RX MEDLINE=78137069; PubMed=416441;

RA Torano A., Putnam F.W.;

RT "Complete amino acid sequence of the alpha 2 heavy chain of a human

RT IGA2 immunoglobulin of the A2m (2) allotype";

RL Proc. Natl. Acad. Sci. U.S.A. 75:968-969(1978).

CC -I- MISCELLANEOUS: THE SEQUENCE OF THE ALPHA-2, A2M(2) ALLOTYPE, C

CC REGION OF THIS MYELOMA PROTEIN IS ALSO GIVEN.

CC -I- SIMILARITY: Contains 1 immunoglobulin-like domain.

DR PIR; A02050; A2HUBU.

DR HSSP; P01789; 1MCP.

DR GO; GO:0005576; C:extracellular; NAS.

DR GO; GO:0003823; F:antigen binding; NAS.

DR GO; GO:0006955; P:immune response; NAS.

DR InterPro; IPR007110; IG-like.

DR InterPro; IPR003596; IG\_v.

DR Pfam; PF00047; Ig; 1.

DR SMART; SM00406; IGV; 1.

DR PROSITE; PS50835; IG\_LIKE; 1.

DR Immunoglobulin V region.

FT DOVAIN 1 111 IG-LIKE.

FT NON TER 115 115

SQ SEQUENCE 115 AA; 12379 MW; 208876A7DF52DCE4 CRC64;

Query Match 48.4%; Score 335; DB 1; Length 115;

Best Local Similarity 57.1%; Pred. No. 2.9e-26;

Matches 68; Conservative 17; Mismatches 28; Indels 6; Gaps 1;

QY	1	QVQLDSGGGLVQAGGSLRLSCAASGIGDIHTMGWYRQTPGKQDVVATIQSGSTNYA	60
DB	1	EVQLVETGGGLIQPGSLRLSCAASGFTVSBHSMVSRQAPGKALZWVSAIYRGTTIYA	60
QY	61	DSVKGRTITSDNTLVYQLNDLKPEDTGYYTNADVPYRISRYLEIWGGQTLTVV	119
DB	61	DSVKGRTITSDSDRBTYVQLMSLRADTAIVYCARDLAAR-----LFGKGTITTV	113

RESULT 2

HV05 CARAU

ID HV05 CARAU STANDARD; PRT; 116 AA.

AC P19181;

DT 01-NOV-1990 (Rel. 16, Created)

DT 01-NOV-1990 (Rel. 16, Last sequence update)

```

DR 15-JUL-1999 (Rel. 38, Last annotation update)
DE Ig heavy chain V region 5A precursor.
OS Carassius auratus (Goldfish).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Carassius.
OX NCBI_TaxID=7957;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88144476; PubMed=3125551;
RA Wilson M.R., Middleton D., Warr G.W.;
RT "Immunoglobulin heavy chain variable region gene evolution: structure
RT and family relationships of two genes and a pseudogene in a teleost
RT fish."
RL Proc. Natl. Acad. Sci. U.S.A. 85:1566-1570(1988).
DR HSP; P01772; 2FB4.
DR InterPro; IPR007110; IG-like.
DR SMART; SM00406; IGV; 1.
DR InterPro; IPR003596; IG_V.
DR Pfam; PF00047; IGV; 1.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS0835; IG LIKE; 1.
DR Immunoglobulin V region; Pyrrolidone carboxylic acid.
KW DOMAIN 1 112 IG-LIKE
FT MOD_RES 1 1 PYRROLIDONE CARBOXYLIC ACID.
FT NON_TER 122 122
SQ SEQUENCE 122 AA; 13668 MW; A42D0F17D252F1C2 CRC64;

Query Match 46.0%; Score 318.5; DB 1; Length 122;
Best Local Similarity 55.8%; Pred. No. 1.3e-24;
Matches 67; Conservative 13; Mismatches 39; Indels 1; Gaps 1;

QY 1 QVQLDSGGGLVQAGGSLRLSCAASGSIQDTHMGWYRQTPGKQRDVVATTQ-SGGSTNY 59
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 QVELVESGGVVPGRSLRLSCAASGFTFSNYAMHWVRQPFKGLWEVAVISYGBBKYY 60
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 60 ADSVKGRTISRNTLNTVYLQNDLKPEDTGYYVWADVRPRTSRYLEIWGGTLTV 119
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 61 ABSVKGRTISRNTLNTVYLQNDLKPEDTGYYVWADVRPRTSRYLEIWGGTLTV 120

RESULT 4
HV3G HUMAN STANDARD; PRT; 119 AA.
ID HV3N HUMAN
AC P01775;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ig heavy chain V-III region LAY.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE.
RX MEDLINE=75046755; PubMed=4139708;
RA Capra J.D., Kehoe J.M.;
RT "Structure of antibodies with shared idiotypic: the complete sequence
RT of the heavy chain variable regions of two immunoglobulin M
RT anti-gamma globulins."
RT Proc. Natl. Acad. Sci. U.S.A. 71:4032-4036(1974).
RL CC -!- MISCELLANEOUS; THIS CHAIN WAS ISOLATED FROM IGM WITH ANTI-GAMMA
CC GLOBULIN ACTIVITY.
CC -!- SIMILARITY: Contains 1 immunoglobulin-like domain.
DR PIR; A02058; M3HULY.
DR HSSP; P01772; 2FB4.
DR GO; GO:0005576; C:extracellular; NAS..
DR GO; GO:0003823; F:antigen binding; NAS.
DR GO; GO:0006955; P:immune response; NAS.
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003596; IG_V.
DR Pfam; PF00047; IGV; 1.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS0835; IG LIKE; 1.
KW Immunoglobulin V region.
FT DOMAIN 1 112 IG-LIKE.
FT NON_TER 119 119
SQ SEQUENCE 119 AA; 12858 MW; D6338098794DCF5E CRC64;

Query Match 45.9%; Score 317.5; DB 1; Length 119;
Best Local Similarity 55.1%; Pred. No. 1.6e-24;
Matches 65; Conservative 15; Mismatches 37; Indels 1; Gaps 1;

QY 2 VQLQDSGGGLVQAGGSLRLSCAASGSIQDTHMGWYRQTPGKQRDVVA-TIQSGSTNYA 60
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 2 VQLLESGGGLVQPGGSLRLSCAASGFTFSASANSVVRQAPFGKLEWAVKVGNDKHYA 61
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 61 DSVKGRFTISRNTLNTVYLQNDLKPEDTGYYVWADVRPRTSRYLEIWGGTLTV 118
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 62 DSVNKGRTISRNTLNTVYLQNDLKPEDTGYYVWADVRPRTSRYLEIWGGTLTV 119

Query Match 45.9%; Score 317.5; DB 1; Length 119;
Best Local Similarity 55.1%; Pred. No. 1.6e-24;
Matches 65; Conservative 15; Mismatches 37; Indels 1; Gaps 1;

QY 2 VQLQDSGGGLVQAGGSLRLSCAASGSIQDTHMGWYRQTPGKQRDVVA-TIQSGSTNYA 60
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 2 VQLLESGGGLVQPGGSLRLSCAASGFTFSASANSVVRQAPFGKLEWAVKVGNDKHYA 61
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 61 DSVKGRFTISRNTLNTVYLQNDLKPEDTGYYVWADVRPRTSRYLEIWGGTLTV 118
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 62 DSVNKGRTISRNTLNTVYLQNDLKPEDTGYYVWADVRPRTSRYLEIWGGTLTV 119

```

```

RESULT 5
HV3U HUMAN
ID HV3U HUMAN STANDARD; PRT; 120 AA.
AC P01782;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ig heavy chain V-III region DOB.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCBI_TaxID=9606;
[1]
RP SEQUENCE.
RX MEDLINE=80020921; PubMed=114209;
RA Steiner L.A., Garcia Pardo A., Margolies M.N.;
RT "Amino acid sequence of the heavy-chain variable region of the
RT crystallizable human myeloma protein Dob.";
RL Biochemistry 18:4058-4080(1979).
[2]
CC -!- MISCELLANEOUS: THIS GAMMA-1 MYELOMA PROTEIN HAS A DELETION IN THE
CC HINGE REGION. THERE ARE NO LIGHT-HEAVY OR INTER-HEAVY CHAIN
CC DISULFIDE BONDS.
CC -!- SIMILARITY: Contains 1 immunoglobulin-like domain.
DR PIR; A90431; GIHUDB.
DR HSP; P01772; 2FB4.
DR GO; GO:0005576; C:extracellular; NAS.
DR GO; GO:0003823; F:antigen binding; NAS.
DR GO; GO:0006955; P:immune response; NAS.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003596; Ig_v.
DR Pfam; PF00047; Ig; 1.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS00835; IG LIKE; 1.
KW Immunoglobulin V region.
FT DOMAIN 1 112 IG-LIKE.
FT MOD_RES 22 96
FT DISULFID 119 119
FT NON_TER 120 120
SQ SEQUENCE 120 AA; 13440 MW; 880DDE307C4B2627 CRC64;

Query Match 45.7%; Score 316.5; DB 1; Length 120;
Best Local Similarity 55.6%; Pred. No. 28-24;
Matches 70; Conservative 12; Mismatches 29; Indels 15; Gaps 3;

QY 1 QVQLQDSGGGLVQAGGSLRLSCAASGSGIDHTMGWYRQTFGKQDVVATIQ-SGGSTNY 59
Db 1 EVQLVESGGDLVQPGSRSLRSCAASGFNFEYHNHMLRQPGKPEWVSITITWNGGSVLY 60

QY 60 ADSVKGRTISRDNLTNTVLYOMNDLKPEDTGVTY-----WNADVPYRTSYLEIWQ 113
Db 61 ADSVKGRTISRDNLTNTVLYOMNDLKPEDTGVTY-----WFDSWGQ 112

QY 114 GTLVTV 119
Db 113 GTLVTV 118

RESULT 6
HV3I HUMAN
ID HV3I HUMAN STANDARD; PRT; 119 AA.
AC P01770;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ig heavy chain V-III region NIE.
OS Homo sapiens (Human).

```

```

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCBI_TaxID=9606;
[1]
RP SEQUENCE.
RX MEDLINE=77070269; PubMed=826475;
RA Ponstingl H., Hilschmann N.;
RT "The rule of antibody structure. The primary structure of a
RT monoclonal IgG1 immunoglobulin (myeloma protein NIE). III. The
RT chymotryptic peptides of the H-chain, alignment of the tryptic
RT peptides and discussion of the complete structure.";
RL Hoppe-Seyler's Z. Physiol. Chem. 357:1515-1540(1976).
[2]
CC -!- MISCELLANEOUS: THIS CHAIN WAS ISOLATED FROM AN IGG1 MYELOMA
CC PROTEIN.
CC -!- SIMILARITY: Contains 1 immunoglobulin-like domain.
DR PIR; A91668; GIHUNI.
DR HSP; P01772; 2FB4.
DR GO; GO:0005576; C:extracellular; NAS.
DR GO; GO:0003823; F:antigen binding; NAS.
DR GO; GO:0006955; P:immune response; NAS.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003596; Ig_v.
DR Pfam; PF00047; Ig; 1.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS00835; IG LIKE; 1.
KW Immunoglobulin V region; Pyrrolidone carboxylic acid.
FT DOMAIN 1 112 IG-LIKE.
FT MOD_RES 22 96
FT DISULFID 119 119
FT NON_TER 119 119
SQ SEQUENCE 119 AA; 13242 MW; C96935A6B55E165B CRC64;

Query Match 45.4%; Score 314; DB 1; Length 119;
Best Local Similarity 57.5%; Pred. No. 3.6e-24;
Matches 69; Conservative 14; Mismatches 33; Indels 4; Gaps 3;

QY 1 QVQLQDSGGGLVQAGGSLRLSCAASGSGIDHTMGWYRQTFGKQDVVATIQ-SGGSTNY 59
Db 1 QVQLVQSGGVVQPGKSLRSLRSCAASGFTFSRYTIHWVQAPGKGLWVAVMSYBGBKH 60

QY 60 ADSVKGRTISRDNLTNTVLYOMNDLKPEDTGVTY-----WNADVPYRTSYLEIWQ 119
Db 61 ADSVKGRTISRDNLTNTVLYOMNDLKPEDTGVTY-----TAMFAHWGQGLTVTV 117

RESULT 7
HV3B HUMAN
ID HV3B HUMAN STANDARD; PRT; 114 AA.
AC P01763;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ig heavy chain V-III region WEA.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCBI_TaxID=9606;
[1]
RP SEQUENCE.
RX MEDLINE=83273707; PubMed=6410398;
RA Coni F., Frangione B.;
RT "Amino acid sequence of the Fv region of a human monoclonal IgM
RT (protein WEA) with antibody activity against 3,4-pyruvylated
RT galactose in Klebsiella polysaccharides K30 and K33.";

```

```

RL Proc. Natl. Acad. Sci. U.S.A. 80:4837-4841 (1983).
CC -1- MISCELLANEOUS: THIS CHAIN WAS OBTAINED FROM A MONOCLONAL ANTIBODY
CC AGAINST 3,4-PYRUVYLATED GALACTOSE AND ISOLATED FROM A PATIENT WITH
CC WALDENSTROM'S MACROGLOBULINEMIA.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.
DR PIR; A02046; M3HUP6.
DR HSP; P01772; 2FB4.
DR GO; GO:0005576; C:extracellular; NAS.
DR GO; GO:0003823; F:antigen binding; NAS.
DR GO; GO:0006955; P:immune response; NAS.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003596; Ig_v.
DR Pfam; PF00047; ig; 1.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS00835; IG_LIKE; 1.
KW Immunoglobulin V region; Pyrolydione carboxylic acid.
FT DOMAIN 1 112 IG-LIKE.
FT MOD_RES 1 1 PYRROLIDONE CARBOXYLIC ACID.
FT NON_TER 114 114
SQ SEQUENCE 114 AA; 12256 MW; D88294FB418A07B7 CRC64;

Query Match 45.3%; Score 313.5; DB 1; Length 114;
Best Local Similarity 56.7%; Pred. No. 3.8e-24;
Matches 68; Conservative 14; Mismatches 29; Indels 9; Gaps 2;

QY 1 QVQLDSGGGLVQAGGSLRLSCAASGIGDIHTMGWYRQTPGKQRDVVATIQSGGST-NY 59
DB 1 QVQLVDSGGGLVPEGGSRLSCASGFTFSANDMNVWVQAPGKGLWLSFTGGSGSTIYY 60

QY 60 ADSVKGRFTISRDNLTNTLYLQMDLKPEDTGVYNNADVRPYRTSRYLEIHWGQTLTV 119
DB 61 ADSVKGRFTISRDNLTNTLYLQMDLKPEDTGVYNNADVRPYRTSRYLEIHWGQTLTV 112

RESULT 8
HV3C_HUMAN STANDARD; PRT; 117 AA.
AC P01764;
DC 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ig heavy chain V-III region VH2 precursor.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=81101090; PubMed=6450418;
RA Mathysse G., Rabbitts T.H.;
RT "Structure and multiplicity of genes for the human immunoglobulin
RT heavy chain variable region."
RL Proc. Natl. Acad. Sci. U.S.A. 77:6561-6565 (1980).
CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; J00236; AAA53516.1; -
DR EMBL; M35415; AAA58735.1; -
DR PIR; A02047; H3H26.
DR PDB; 1HOU; 23-DEC-99.
DR Genew; HGNC:5545; IGHV@.
DR GO; GO:0005576; C:extracellular; NAS.
DR GO; GO:0003823; F:antigen binding; NAS.
DR GO; GO:0006955; P:immune response; NAS.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003596; Ig_v.

```

```

DR Pfam; PF00047; ig; 1.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS00835; IG_LIKE; 1.
KW Immunoglobulin V region; Signal; 3D-structure.
FT SIGNAL 1 19 IG HEAVY CHAIN V-III REGION VH26.
FT CHAIN 20 117 IG-LIKE.
FT DOMAIN 20 >117 IG-LIKE.
FT NON_TER 117 117
SQ SEQUENCE 117 AA; 12582 MW; E826733F1A3CB0F1 CRC64;

Query Match 45.2%; Score 312.5; DB 1; Length 117;
Best Local Similarity 67.4%; Pred. No. 4.9e-24;
Matches 64; Conservative 8; Mismatches 22; Indels 1; Gaps 1;

QY 1 QVQLDSGGGLVQAGGSLRLSCAASGIGDIHTMGWYRQTPGKQRDVVATIQSGGSTNY 59
DB 20 EVQLLESGGGLVQPGGSLRLSCAASGFTFSVANSVWVQAPGKGLWVSAISGSGSTYY 79

QY 60 ADSVKGRFTISRDNLTNTLYLQMDLKPEDTGVY 94
DB 80 GDSVKGRFTISRDNLTNTLYLQMDLKPEDTGVY 114

RESULT 9
HV3M_HUMAN STANDARD; PRT; 119 AA.
AC P01774;
DC 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ig heavy chain V-III region POM.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE.
RX MEDLINE=75046755; PubMed=4139708;
RA Capra J.D., Kehoe J.M.;
RT "Structure of antibodies with shared idiotype: the complete sequence
RT of the heavy chain variable regions of two immunoglobulin M
RT anti-gamma globulins."
RL Proc. Natl. Acad. Sci. U.S.A. 71:4032-4036 (1974).
CC -1- MISCELLANEOUS: THIS CHAIN WAS ISOLATED FROM IGM WITH ANTI-GAMMA
CC GLOBULIN ACTIVITY.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.
DR PIR; A02057; M3HUP6.
DR HSP; P01772; 2FB4.
DR GO; GO:0005576; C:extracellular; NAS.
DR GO; GO:0003823; F:antigen binding; NAS.
DR GO; GO:0006955; P:immune response; NAS.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003596; Ig_v.
DR Pfam; PF00047; ig; 1.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS00835; IG_LIKE; 1.
KW Immunoglobulin V region.
FT DOMAIN 1 112 IG-LIKE.
FT VARIANT 54 54 N -> D (PROBABLY DUE TO DEAMINATION
FT DURING ISOLATION).
FT FTID=VAR_003966.
FT NON_TER 119 119
SQ SEQUENCE 119 AA; 12953 MW; 2E018AF4DCB2610 CRC64;

Query Match 45.2%; Score 312.5; DB 1; Length 119;
Best Local Similarity 53.8%; Pred. No. 5e-24;
Matches 64; Conservative 17; Mismatches 37; Indels 1; Gaps 1;

QY 1 QVQLDSGGGLVQAGGSLRLSCAASGIGDIHTMGWYRQTPGKQRDVVATIQSGGSTNY 59
DB 1 EVQLLESGGGLVQPGGSLRLSCAASGFTFSVANSVWVQAPGKGLWVSAISGSGSTYY 60

QY 60 ADSVKGRFTISRDNLTNTLYLQMDLKPEDTGVYNNADVRPYRTSRYLEIHWGQTLTV 118

```





```

Db      111 LVTV 114
|||||
RESULT 12
HV33 MOUSE
ID HV33 MOUSE STANDARD; PRT; 115 AA.
AC P01802;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ig heavy chain V-III region W3082.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE.
RX MEDLINE=82099361; PubMed=6798111;
RA Johnson N., Sankard J., Paul L.;
RT "The complete V domain amino acid sequences of two myeloma inulin-
binding proteins."
RT J. Immunol. 128:302-307(1982).
CC -!- MISCELLANEOUS: THIS CHAIN WAS ISOLATED FROM A MYELOMA PROTEIN THAT
BINDS INULIN.
CC -!- SIMILARITY: Contains 1 immunoglobulin-like domain.
DR PIR; D92811; AVMS82.
DR HSSP; P01810; 2FBJ.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003596; Ig_v.
DR Pfam; PF00047; Ig; 1.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS00835; IG LIKE; 1.
KW Immunoglobulin V region.
FT DOMAIN 1 114 IG-LIKE.
FT DISULFID 22 98 BY SIMILARITY.
FT NON_TER 115 115
SQ SEQUENCE 115 AA; 12887 MW; 984517648C121C5A CRC64;

Query Match 44.1%; Score 305; DB 1; Length 115;
Best Local Similarity 51.6%; Pred. No. 2,6e-23;
Matches 63; Conservative 21; Mismatches 26; Indels 12; Gaps 2;

QY 1 QVQLDSGGGLVQAGSRLSCAASGSDGHTMHWYRQTPKQRDVVATIQSGST 57
Db 1 EVKLESGGGLVQPGSMKLSVAGSFYSNYMMWVRQSPKGLWVAEIRLASHNAT 60
QY 58 NYADSVKGRFTISRDNLTNTVYLNQNDLKPEDTGVYVYNADVRPRTSRYLEINGQGLV 117
Db 61 HVAESVKGRTISRDSKSSVYLRWNNLRPDTGIYCTTG-----PAYWGQGLV 111
QY 118 TV 119
Db 112 TV 113

RESULT 13
HV3E HUMAN
ID HV3E HUMAN STANDARD; PRT; 120 AA.
AC P01766;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ig heavy chain V-III region BRO.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE.
RX MEDLINE=77117674; PubMed=65324;
RA Capra J.D., Hopper J.E.;
RT "Comparative studies on monotypic IgM lambda and IgG kappa from an

```

```

RT individual patient. III. The complete amino acid sequence of the VH
region of the IgM paraprotein."
RL Immunohistochemistry 13:995-999(1976).
CC -!- MISCELLANEOUS: THIS CHAIN WAS OBTAINED FROM IGM ISOLATED FROM THE
SERUM OF A PATIENT WITH MALIGNANT LYMPHOMA OF THE WALDENSTROM
TYPE.
CC -!- SIMILARITY: Contains 1 immunoglobulin-like domain.
DR PIR; A02049; M3HUBW.
DR HSSP; P01772; 2FB4.
DR GO; GO:0005576; C:extracellular; NAS.
DR GO; GO:0003823; F:antigen binding; NAS.
DR GO; GO:0006955; P:immune response; NAS.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003596; Ig_v.
DR Pfam; PF00047; Ig; 1.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS00835; IG LIKE; 1.
KW Immunoglobulin V region.
FT DOMAIN 1 111 IG-LIKE.
FT NON_TER 120 120
SQ SEQUENCE 120 AA; 13227 MW; D3F0428F7C2E6410 CRC64;

Query Match 43.9%; Score 304; DB 1; Length 120;
Best Local Similarity 54.2%; Pred. No. 3,5e-23;
Matches 65; Conservative 13; Mismatches 38; Indels 4; Gaps 1;

QY 1 QVQLDSGGGLVQAGSRLSCAASGSDGHTMHWYRQTPKQRDVVATIQSGSTNYA 60
Db 1 EVQLVESGGGLVQPGGSLRLSCAASGFTFSYNNMWVRQVTKGLEWVAISATGQYYA 60
QY 61 DSVKGRFTISRDNLTNTVYLNQNDLKPEDTGVYVYNAD---VRPRTSRYLEINGQGLT 116
Db 61 DSVKGRFTISRDNLTNTVYLNQNDLKPEDTGVYVYNAD---VRPRTSRYLEINGQGLT 120

RESULT 14
HV16 MOUSE
ID HV16 MOUSE STANDARD; PRT; 136 AA.
AC P01783;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ig heavy chain V region MOPC 21 precursor (Fragment).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=81234548; PubMed=6788376;
RA Bothwell A.L.M., Paskind M., Rath M., Imanishi-Kari T., Rajewsky K.,
RA Baltimore D.;
RT "Heavy chain variable region contribution to the NPb family of
antibodies: somatic mutation evident in a gamma 2a variable region."
RL Cell 24:625-637(1981).
RN [2]
RP SEQUENCE OF 17-136.
RX MEDLINE=77100368; PubMed=401950;
RA Adetugbo K., Milstein C., Secher D.S.;
RT "Molecular analysis of spontaneous somatic mutants."
RL Nature 265:299-304(1977).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC EMBL; J00522; AAD15290.1; -.
DR PIR; E90809; G1MS21.
DR PDB; 1IGC; 03-JUN-95.

```

```

DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003596; Ig_v.
DR Pfam; PF00047; Ig; 1.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS00835; IG LIKE; 1.
KW Immunoglobulin V region; Signal; 3D-structure.
FT NON_TER 1
FT SIGNAL 16
FT CHAIN 17 IG HEAVY CHAIN V REGION MOPC 21.
FT DOMAIN 115 D SEGMENT.
FT DOMAIN 119 JH4 SEGMENT.
FT DOMAIN 120 136
FT DISULFID 38 112
FT CONFLICT 75 78
FT CONFLICT 89 90
FT CONFLICT 115 115
FT CONFLICT 120 120
FT CONFLICT 136 136
FT NON_TER 136
SQ SEQUENCE 136 AA; 15071 MW; 2276A98DBDF7016 CRC64;

Query Match 43.8%; Score 303.5; DB 1; Length 136;
Best Local Similarity 57.4%; Pred. No. 4.5e-23;
Matches 70; Conservative 9; Mismatches 34; Indels 9; Gaps 4;

QY 2 VQLQDSGGGLVQAGGSLRLSCAASGSIQDIHTMGWVRQTPGKQRDVVATIQSGST-NYA 60
Db 18 VQLVESGGGLVQPGSRKLSCAASGFTFSFGMHVYRQAEKGLWVAIYSSGSSTLHYA 77

QY 61 DSVKGRFTISRDNTLVYLNQNDLKPEDTGVY---WNADVPYRTSRYLEIWGGQTLV 117
Db 78 DTVKGRFTISRDNPKNLTFLQMTLSRSEDYATYCARWGN--YFYANDY---WGGQTSV 132

QY 118 TV 119
Db 133 TV 134

RESULT 15
HV32 MOUSE
ID HV32_MOUSE STANDARD; PRT; 115 AA.
AC P01801;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ig heavy chain V-III region J606.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_TaxID=10090;
RN [1]
RP SEQUENCE.
RX MEDLINE=82099361; PubMed=6798111;
RA Johnson N., Slankard J., Paul L., Hood L.;
RT "The complete V domain amino acid sequences of two myeloma inulin-
binding proteins.";
RL J. Immunol. 128:302-307(1992).
CC -!- MISCELLANEOUS: THIS CHAIN WAS ISOLATED FROM A MYELOMA PROTEIN THAT
CC BINDS INULIN.
CC -!- SIMILARITY: Contains 1 immunoglobulin-like domain.
DR PIR; C92811; AVMS06.
DR HSSP; P01789; IMCP.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003596; Ig_v.
DR Pfam; PF00047; Ig; 1.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS00835; IG LIKE; 1.
KW Immunoglobulin V region.
FT DOMAIN 1 114 IG-LIKE.
FT DISULFID 22 98 BY SIMILARITY.
FT NON_TER 115
SQ SEQUENCE 115 AA; 12810 MW; B67AD6638A121A5F CRC64;

Query Match 43.8%; Score 303; DB 1; Length 115;
Best Local Similarity 51.6%; Pred. No. 4.2e-23;

```

```

Matches 63; Conservative 21; Mismatches 26; Indels 12; Gaps 2;
QY 1 QVQLDSSGGLVQAGGSLRLSCAASGSIQDIHTMGWVRQTPGKQRDVVATIQ---SGST 57
Db 1 EVKLEESGGGLVQPGGSKLSKVASGFTFSNYMMWVRQSPKGLWVAEIRLKSNNYAT 60

QY 58 NYADSVKGRFTISRDNTLVYLNQNDLKPEDTGVYWNADVPYRTSRYLEIWGGQTLV 117
Db 61 HYAESVGRFTISRDSSKSSVYLQNNLRADETGLIYYCTTG-----FAYWGGQTLV 111

QY 118 TV 119
Db 112 TV 113

```

Search completed: October 5, 2004, 08:01:57  
Job time : 11.6358 secs

This Page Blank (uspb)

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2004, 07:34:11 ; Search time 68.8:17 Seconds  
(without alignments)  
596.081 Million cell updates/sec

Title: US-09-805-290A-26

Perfect score: 692  
Sequence: 1 QVQLQDSGGGLVAGGSLRL.....WGQSTLVTVFLPEKTPKPOP 130

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL 25:\*

- 1: sp archaea:\*
- 2: sp bacteria:\*
- 3: sp fungi:\*
- 4: sp human:\*
- 5: sp invertebrate:\*
- 6: sp mammal:\*
- 7: sp mhc:\*
- 8: sp organelle:\*
- 9: sp phage:\*
- 10: sp plant:\*
- 11: sp rodent:\*
- 12: sp virus:\*
- 13: sp vertebrate:\*
- 14: sp unclassified:\*
- 15: sp\_rvirus:\*
- 16: sp\_bacteriap:\*
- 17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	358	51.7	597	4 Q96B99	Q96bb9 homo sapien
2	349.5	50.5	118	4 Q9UL72	Q9ul72 homo sapien
3	346.5	50.1	499	4 Q8N5K4	Q8n5k4 homo sapien
4	336.5	48.6	613	4 Q8WUK1	Q8wuk1 homo sapien
5	335	48.4	113	4 Q9UL90	Q9ul90 homo sapien
6	334.5	48.3	487	11 Q99KA4	Q99ka4 mus musculus
7	332	48.0	121	4 Q9UL71	Q9ul71 homo sapien
8	331.5	47.9	112	4 Q9HCC1	Q9hcc1 homo sapien
9	326.5	47.2	118	4 Q9UL91	Q9ul91 homo sapien
10	326	47.1	119	11 Q92OE7	Q92oe7 mus musculus
11	324	46.8	480	11 Q91XE1	Q9ixei mus musculus
12	324	46.8	486	11 Q91Z07	Q91z07 mus musculus
13	324	46.8	573	4 Q8WU38	Q8w38 homo sapien
14	319.5	46.2	147	4 Q9Y509	Q9y509 homo sapien
15	317	45.8	131	4 Q9UL88	Q9ul88 homo sapien
16	317	45.8	479	11 Q91WP5	Q91wp5 mus musculus

17	315.5	45.6	122	4 Q9UL84	Q9ul84 homo sapien
18	314	45.4	487	11 Q80Z17	Q80z17 mus musculus
19	313	45.2	494	4 Q96K68	Q96k68 homo sapien
20	313	45.2	521	4 Q8N4Y9	Q8n4y9 homo sapien
21	312	45.1	116	4 Q9UL93	Q9ul93 homo sapien
22	304.5	44.0	469	11 Q8R3V9	Q8r3v9 mus musculus
23	302.5	43.7	437	11 Q9R1A4	Q9ria4 mus musculus
24	302.5	43.7	473	11 Q9N1Z05	Q9n1z05 mus musculus
25	302.5	43.7	493	4 Q8NCL6	Q8nc16 homo sapien
26	296.5	42.8	95	4 Q9ULB6	Q9ulb6 homo sapien
27	290.5	42.0	298	11 Q9QYF0	Q9qyf0 mus musculus
28	289	41.8	470	4 Q7Z5W1	Q7z5w1 homo sapien
29	275.5	39.8	124	6 Q9N0W6	Q9n0w6 cryctolagus
30	275.5	39.8	484	11 Q8VEA0	Q8vea0 mus musculus
31	273.5	39.5	124	6 Q9N0W4	Q9n0w4 cryctolagus
32	271.5	39.2	479	11 Q7TMR4	Q7tmt4 mus musculus
33	254	36.7	482	4 Q7Z351	Q7z351 homo sapien
34	253	36.6	104	4 Q9UL87	Q9ul87 homo sapien
35	245.5	35.5	124	4 Q9UL92	Q9ul92 homo sapien
36	245	35.4	112	4 Q9UGP3	Q9ugp3 homo sapien
37	241.5	34.9	482	11 Q9IX32	Q9ix32 mus musculus
38	241	34.8	614	11 Q7TMT6	Q7tmt6 mus musculus
39	232	33.5	119	4 Q9UL73	Q9ul73 homo sapien
40	230	33.2	125	4 Q9UL95	Q9ul95 homo sapien
41	229.5	33.2	168	11 Q8VDC9	Q8vdc9 mus musculus
42	229.5	33.2	613	11 Q8VCX7	Q8vcx7 mus musculus
43	228	32.9	159	4 Q8QSO0	Q8qso0 homo sapien
44	228	32.9	614	4 Q96GA6	Q96ga6 homo sapien
45	227.5	32.9	121	11 Q99NG4	Q99ng4 mus musculus

ALIGNMENTS

RESULT 1

Q96BB9 ID Q96BB9 PRELIMINARY; PRT; 597 AA.  
AC Q96BB9; DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Hypothetical protein.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
CX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=B-cell;  
RA Strausberg R.;  
RL Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.  
DR EMBL; BC015760; AAH15760.1; -  
DR InterPro; IPR007110; IG-like.  
DR InterPro; IPR003006; IG\_MHC.  
DR InterPro; IPR003596; IG\_V.  
DR Pfam; PF00047; IG; 5.  
DR SMART; SM00406; IGV; 1.  
DR PROSITE; PS00835; IG LIKE; 5.  
DR PROSITE; PS00290; IG\_MHC; 3.  
KW Hypothetical protein.  
SQ SEQUENCE 597 AA; 65039 MW; 4FCA3AD8ECE263D9 CRC64;

Query Match 51.7%; Score 358; DB 4; Length 597;  
Best Local Similarity 63.4%; Pred. NO. 1.9e-25;  
Matches 78; Conservative 10; Mismatches 31; Indels 4; Gaps 3;

Qy	1	QVQLQDSGGGLVAGGSLRLCAAGSICDIHTMGWYRTPGKQRDVATIQ-SGSGTMY	59
Db	20	EVLLSGGGLVQPGGSLRLCAAGSIFSSVAMWVRQAPKGLWVAISGSGSTYY	79
Qy	60	ADSVKGRFTISRDNTLVYLNKLPEDTGYVYWNADVRYPTS-RYL-EINGQGT	116
Db	80	ADSVKGRFTISRDNTLVYLNKLPEDTGYVYWNADVRYPTS-RYL-EINGQGT	139

```

QY 117 VTV 119
DB 140 VTV 142

RESULT 2
Q9UL72 PRELIMINARY; PRT; 118 AA.
AC Q9UL72;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Myosin-reactive immunoglobulin heavy chain variable region
DE (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98277139; PubMed=9614934;
RA Wu X., Liu B., van der Merwe P.B., Kalis N.N., Berney S.M.,
RT "Myosin-reactive autoantibodies in rheumatic carditis and normal
RT fetus."
RL Clin. Immunol. Immunopathol. 87:184-192(1998).
DR EMBL; AF035042; AAD56278.1; -.
DR PIR; S21205; S21205.
DR HSSP; P01772; 2FB4.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003596; Ig_v.
DR SMART; PF00047; IG; 1.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS00835; IG_LIKE; 1.
FT NON_TER 1
FT NON_TER 118
SQ SEQUENCE 118 AA; 12872 MW; B4DIA5944B2D5CCA CRC64;

Query Match 50.5%; Score 349.5; DB 4; Length 118;
Best Local Similarity 61.7%; Pred. No. 1.5e-25;
Matches 74; Conservative 11; Mismatches 30; Indels 5; Gaps 2;

QY 1 QVQLQDSGGGLVQAGSLRLSCAASGSIQDIHTMGWYRQTPGKQRDVVATIQSGGSTNYA 60
DB 1 EVQLVESGGGVLPQGGSLRLSCAASGFTVSSNNMWRQAPGKGLSVSVTSGGSYYA 60

QY 61 DSVKGRFTISRDTLTNTVYLNQNDLKPEDTGVYYNADVRPYRTSKYL-ETWGGQTLTV 119
DB 61 DSVKGRFTISRDNKNTLYLNQNSLRVEDTALYYCARDPTKYCSGGSLGYMDVWGK 116

RESULT 3
Q8N5K4 PRELIMINARY; PRT; 499 AA.
AC Q8N5K4;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=Blood;
RA Strausberg R.;
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC032249; AAH32249.1; -.
DR InterPro; IPR003599; IG.
DR InterPro; IPR007110; IG_LIKE.
DR InterPro; IPR003597; IG_c1.

```

```

DR InterPro; IPR003006; IG_MHC.
DR InterPro; IPR003596; IG_v.
DR Pfam; PF00047; IG; 4.
DR SMART; SM00409; IG; 4.
DR SMART; SM00407; IGV; 2.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS00835; IG_LIKE; 4.
DR PROSITE; PS00290; IG_MHC; 1.
KW Hypothetical protein.
SQ SEQUENCE 499 AA; 53376 MW; 93A5C89582054F32 CRC64;

Query Match 50.1%; Score 346.5; DB 4; Length 499;
Best Local Similarity 54.0%; Pred. No. 1.9e-24;
Matches 74; Conservative 18; Mismatches 38; Indels 7; Gaps 3;

QY 1 QVQLQDSGGGLVQAGSLRLSCAASGSIQDIHTMGWYRQTPGKQRDVVATIQSGGSTNY 59
DB 20 EVQLVESGGGVLPQGGSLRLSCATSGFTFDDSGASWVRQAPGKGLWVSSINWNGSTNY 79

QY 60 ADSVKGRFTISRDTLTNTVYLNQNDLKPEDTGVYYNADVRPYRT-----SRYLEIWGG 114
DB 80 ADSVKGRFTISRDNKNSLYLNQNSLRVEDTALYYCARDPTKYCSGGSLGYMDVWGK 139

QY 115 TLVTV-FLEPKTPKQP 130
DB 140 TTVTVSSASPTSPKVP 156

RESULT 4
Q8WUK1 PRELIMINARY; PRT; 613 AA.
AC Q8WUK1;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=Tonsil;
RA Strausberg R.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC020240; AAH20240.1; -.
DR PIR; P0120; P0120.
DR PIR; S15590; S15590.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003006; IG_MHC.
DR InterPro; IPR003596; IG_v.
DR Pfam; PF00047; IG; 5.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS00835; IG_LIKE; 5.
DR PROSITE; PS00290; IG_MHC; 3.
KW Hypothetical protein.
SQ SEQUENCE 613 AA; 6075950671E315 CRC64;

Query Match 48.6%; Score 336.5; DB 4; Length 613;
Best Local Similarity 58.9%; Pred. No. 2.2e-23;
Matches 73; Conservative 11; Mismatches 29; Indels 11; Gaps 3;

QY 1 QVQLQDSGGGLVQAGSLRLSCAASGSIQDIHTMGWYRQTPGKQRDVVATIQSGGSTN-Y 59
DB 20 QVQLVESGGGVLPQGGSLRLSCATSGFTFSSYGMHWVRQAPGKGLWVAIVDGSNKY 79

QY 60 ADSVKGRFTISRDTLTNTVYLNQNDLKPEDTGVYY-----WNADVRPYRTSRYLEIWGG 115
DB 80 ADSVKGRFTISRDNKNTLYLNQNSLRVEDTALYYCAKDWSEGVETF-----DIWGGQT 133

QY 116 LMTV 119
DB 134 MVTV 137

```

```
RESULT 5
Q9JUL90
ID Q9JUL90 PRELIMINARY; PRT; 113 AA.
AC Q9JUL90;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Myosin-reactive immunoglobulin heavy chain variable region
DE (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98277139; PubMed=9614934;
RA Wu X., Liu B., Van der Merwe P.L., Kalis N.N., Berney S.M.,
RA Young D.C.;
RT "Myosin-reactive autoantibodies in rheumatic carditis and normal
RT fetus.";
RL Clin. Immunol. Immunopathol. 87:184-192(1998).
RL EMBL; AF035024; AAD56260.1; -.
DR HSSP; P01772; 2FB4.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003596; Ig_v.
DR Pfam; PF00047; Ig; 1.
DR SMART; SM00406; Ig; 1.
DR PROSITE; PS50835; IG_LIKE; 1.
FT NON_TER 1
FT NON_TER 113
SQ SEQUENCE 113 AA; 12437 MW; ED57FDD19086D07F CRC64;

Query Match 48.4%; Score 335; DB 4; Length 113;
Best Local Similarity 60.8%; Pred. No. 3.5e-24;
Matches 73; Conservative 9; Mismatches 28; Indels 10; Gaps 2;

QY 1 QVQLQDSGGGLVQAGGSLRLSCAASGSIQDHTHTMGWYRQTPGKQRDVWATIQSGGSTN-Y 59
DB 1 EVQLVESGGGVQPGGSLRLSCAASGTFSSYGVHWVRQAPGKLEWVAIRYDGSKYY 60
QY 60 ADSVKGRFTISRNTLTNTVYQMNDLKPEDTGVVYWNADV--RPYRTSKYLEIMQGGLTV 119
DB 61 ADSVKGRFTISRDNKNTLYLQMSLRADTAIVYCAK-----LNYWQGGTLTV 111

RESULT 6
Q99KA4
ID Q99KA4 PRELIMINARY; PRT; 487 AA.
AC Q99KA4;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
RL EMBL; BC004786; AA04786.1; -.
DR HSSP; P01810; 2FBJ.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR003596; Ig_v.
DR Pfam; PF00047; Ig; 4.
DR SMART; SM00406; Ig; 1.
DR PROSITE; PS50835; IG_LIKE; 4.
DR PROSITE; PS00290; IG_MHC; 2.
KW Hypothetical protein.
```

```
SQL SEQUENCE 487 AA; 52554 MW; 7DC8E96DB333077B CRC64;

Query Match 48.3%; Score 334.5; DB 11; Length 487;
Best Local Similarity 56.8%; Pred. No. 2.5e-23;
Matches 71; Conservative 14; Mismatches 37; Indels 3; Gaps 2;

QY 1 QVQLQDSGGGLVQAGGSLRLSCAASGSIQDHTHTMGWYRQTPGKQRDVWATIQSGGS-TNY 59
DB 20 EVQLVESGGGLVQPGGSLRLSCAASGTFSSYGVHWVRQTPGKLEWVATISDGGSYTY 79
QY 60 ADSVKGRFTISRNTLTNTVYQMNDLKPEDTGVVYWNADV--RPYRTSKYLEIMQGGLTV 117
DB 80 PNVKGRFTISRDNKNTLYLQMSLRADTAIVYCAKDMGSPYGGYSRFDYWGQGTI 139
QY 118 TVFLE 122
DB 140 TVSSE 144

RESULT 7
Q9JUL71
ID Q9JUL71 PRELIMINARY; PRT; 121 AA.
AC Q9JUL71;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Myosin-reactive immunoglobulin heavy chain variable region
DE (fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98277139; PubMed=9614934;
RA Wu X., Liu B., Van der Merwe P.L., Kalis N.N., Berney S.M.,
RA Young D.C.;
RT "Myosin-reactive autoantibodies in rheumatic carditis and normal
RT fetus.";
RL Clin. Immunol. Immunopathol. 87:184-192(1998).
RL EMBL; AF035043; AAD56279.1; -.
DR HSSP; P01772; 2FB4.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003596; Ig_v.
DR Pfam; PF00047; Ig; 1.
DR SMART; SM00406; Ig; 1.
DR PROSITE; PS50835; IG_LIKE; 1.
FT NON_TER 1
FT NON_TER 121
SQ SEQUENCE 121 AA; 13154 MW; 2F045CCFASD50736 CRC64;

Query Match 48.0%; Score 332; DB 4; Length 121;
Best Local Similarity 59.0%; Pred. No. 7.3e-24;
Matches 72; Conservative 15; Mismatches 29; Indels 6; Gaps 3;

QY 1 QVQLQDSGGGLVQAGGSLRLSCAASGSIQDHTHTMGWYRQTPGKQRDVWATIQ-SGGSTNY 59
DB 1 EVQLVESGGGVQPGGSLRLSCAASGTFDGVYAMHWVRQAPGKLEWVSLISGDSSTYV 60
QY 60 ADSVKGRFTISRNTLTNTVYQMNDLKPEDTGVVYWNADV--RPYRTSKYLEIMQGGLTV 117
DB 61 ADSVKGRFTISRDNKNTLYLQMSLRADTAIVYCAK-----AKGKVTIYDRFDINGQGTIV 117
QY 118 TV 119
DB 118 TV 119

RESULT 8
Q9HCC1
ID Q9HCC1 PRELIMINARY; PRT; 112 AA.
AC Q9HCC1;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
```

```

DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DE Single chain Fv (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Kikuchi M., Takeda C., Tsujimoto Y., Asada S., Nagata K.;
RT "An antibody fragment2A3 specific for native lysozyme :Isolation from a
RL human synthetic phage display library and characterization.";
RL Submitted (Oct-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB049915; BAB16829.1; -.
DR HSSP; P01772; 2FB4.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003596; Ig_v.
DR Pfam; PF00047; Ig; 1.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS50835; IG LIKE; 1.
FT NON_TER 1
FT NON_TER 112
SQ SEQUENCE 112 AA; 12243 MW; 24FLA45EC3B84788 CRC64;

Query Match 47.9%; Score 331.5; DB 4; Length 112;
Best Local Similarity 59.7%; Pred. No. 7.4e-24;
Matches 71; Conservative 13; Mismatches 26; Indels 9; Gaps 3;

QY 1 QVOLODSGGGLVQAGSLRLSCAASGSIIDHTMGWYRQTPGKQRDVATIQ-SGGSTNY 59
DB 1 EVLVESGGGLVQPGGSLRLSCAASGFTFSYSSMMWVRQAPGKGLEWWSYISSTIIYY 60
QY 60 ADSVKGRFTISRDNLTNTVYLNQNDLKPEDTGVYYW-NADVRPYRTSRY-LEIWGGQTLV 117
DB 61 ADSVKGRFTISRDNKNSLYLNQNSLRADTAVY-----CARRALDYWGQGTLLV 112

RESULT 9
Q9UL91 PRELIMINARY; PRT; 118 AA.
AC Q9UL91;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Myosin-reactive immunoglobulin heavy chain variable region (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98277139; PubMed=9614934;
RA Wu X., Liu B., Van der Werwe P.L., Kalis N.N., Berney S.M.,
RT "Myosin-reactive autoantibodies in rheumatic carditis and normal
RT fetus.";
RL Clin. Immunol. Immunopathol. 87:184-192 (1998).
DR EMBL; AF035023; AAD56259.1; -.
DR PIR; S21205; S21205.
DR HSSP; P01772; 2FB4.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003596; Ig_v.
DR Pfam; PF00047; Ig; 1.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS50835; IG LIKE; 1.
FT NON_TER 1
FT NON_TER 118
SQ SEQUENCE 118 AA; 12843 MW; D0633949F2AC149D CRC64;

Query Match 47.2%; Score 326.5; DB 4; Length 118;
Best Local Similarity 59.3%; Pred. No. 2.3e-23;
Matches 73; Conservative 12; Mismatches 31; Indels 7; Gaps 3;

```

```

QY 1 QVOLODSGGGLVQAGSLRLSCAASGSIIDHTMGWYRQTPGKQRDVATIQSGGST-NY 59
DB 1 EVLVESGGGLVQPGGSLRLSCAASGFTFSYSSMMWVRQAPGKGLEWWSYISSTIIYY 60
QY 60 ADSVKGRFTISRDNLTNTVYLNQNDLKPEDTGVYYW-NADVRPYRTSRY-LEIWGGQTLV 118
DB 61 ADSVKGRFTISRDNKNSLYLNQNSLRADTAVY-CARGD-----SSEAFDIWGQGTMTV 115

QY 119 VFL 121
DB 116 VSL 118

RESULT 10
Q920E7 PRELIMINARY; PRT; 119 AA.
AC Q920E7;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Pterin-mimicking anti-idiotope heavy chain variable region (Fragment).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Atkin J.D., Iape A., Jennings I.G., Horaitis O., Cotton R.G.H.;
RT "Definition of the Idiotope of Pterin-Mimicking Antibodies Expressed
RL in Mammalian Cells.";
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF307937; AAL09421.1; -.
DR PIR; C25913; C25913.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003596; Ig_v.
DR Pfam; PF00047; Ig; 1.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS50835; IG LIKE; 1.
FT NON_TER 1
FT NON_TER 119
SQ SEQUENCE 119 AA; 13025 MW; F6E904044381CA7C CRC64;

Query Match 47.1%; Score 326; DB 11; Length 119;
Best Local Similarity 59.3%; Pred. No. 2.6e-23;
Matches 73; Conservative 10; Mismatches 30; Indels 10; Gaps 3;

QY 1 QVOLODSGGGLVQAGSLRLSCAASGSIIDHTMGWYRQTPGKQRDVATIQSGGS-TNY 59
DB 1 EVLVESGGGLVQPGGSLRLSCAASGFTFSYSSMMWVRQTPGKLEWVATISGGSYYTY 60
QY 60 ADSVKGRFTISRDNLTNTVYLNQNDLKPEDTGVYYW-NADVRPYRTSRY---LEIWGGQTL 116
DB 61 PDSVKGRFTISRDNKNTLYLNQNSLRADTAVY-C-----ARHGDDYDVGAYWGQGTLL 114

QY 117 VTV 119
DB 115 VTV 117

RESULT 11
Q91XE1 PRELIMINARY; PRT; 480 AA.
AC Q91XE1;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein (Fragment).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;

```



```
RN      SEQUENCE FROM N.A.
RP      TISSUE=Colon;
RA      Strausberg R.;
RL      Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR      EMBL; BC010798; AAH10798.1; -.
DR      InterPro; IPR007110; Ig-like.
DR      InterPro; IPR003006; IG_MHC.
DR      InterPro; IPR003596; Ig_v.
DR      Pfam; PF00047; Ig; 4.
DR      SMART; SM00406; IGV; 1.
DR      PROSITE; PS00835; IG_LIKE; 4.
DR      PROSITE; PS00290; IG_MHC; 2.
KW      Hypothetical protein.
FT      NON TER
SQ      SEQUENCE 480 AA; 51936 MW; 20B9234BEF2B41ED CRC64;

Query Match          46.8%; Score 324; DB 11; Length 480;
Best Local Similarity 55.8%; Pred. No. 2.4e-22;
Matches 72; Conservative 10; Mismatches 29; Indels 18; Gaps 3;

QY      2 VQLQDSGGGLVQAGGSLRLSCAASGSIQDHTMGWYRQTPGKQDVVATI-QSGGSTNYA 60
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      61 DSVKGRFTISRDNTLNTVYLQMDLKPEDTGVY-----YVNADVPRPYTSRYLEIWGQ 114
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      80 DSMKGRFTISRDNAQNTVLIQMTSLNSEDYAVYCTRGDYW-----YFDVWGAG 128
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      115 TLTVTFLEP 123
Db      129 TTVTSSEP 137

RESULT 12
Q91Z07      PRELIMINARY;      PRT;      486 AA.
ID      Q91Z07
AC      Q91Z07
DT      01-DEC-2001 (TrEMBLrel. 19, Created)
DT      01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT      01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE      Hypothetical protein.
OS      Mus musculus (Mouse).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX      NCBI_TaxID=10090;
RN      [1]
RP      SEQUENCE FROM N.A.
RA      Strausberg R.;
RL      Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR      EMBL; BC010324; AAH10324.1; -.
DR      InterPro; IPR007110; Ig-like.
DR      InterPro; IPR003006; IG_MHC.
DR      Pfam; PF00047; Ig; 4.
DR      SMART; SM00406; IGV; 1.
DR      PROSITE; PS00835; IG_LIKE; 4.
DR      PROSITE; PS00290; IG_MHC; 2.
KW      Hypothetical protein.
SQ      SEQUENCE 486 AA; 52662 MW; 4FEF835125DA870B CRC64;

Query Match          46.8%; Score 324; DB 11; Length 486;
Best Local Similarity 52.8%; Pred. No. 2.5e-22;
Matches 67; Conservative 17; Mismatches 35; Indels 8; Gaps 2;

QY      1 VQLQDSGGGLVQAGGSLRLSCAASGSIQDHTMGWYRQTPGKQDVVATI-QSGGSTNYA 60
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      20 EVHLVESGGGLVQPGKSLKSCVVSFGSFTSYDMSWVRQTPERLEWVAITSGGNTYYP 79
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      61 DSVKGRFTISRDNTLNTVYLQMDLKPEDTGVYVNADVPRPYTSRYLEIWGQGT 115
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      80 DNVKGRFTISRDNAKYLILQMSLSKSEDYAVYCTRGDYW-----YRPEIPIYYSYFDSWGQGT 136
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RN      SEQUENCE FROM N.A.
RP      TISSUE=Colon;
RA      Strausberg R.;
RL      Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR      EMBL; BC010324; AAH10324.1; -.
DR      InterPro; IPR007110; Ig-like.
DR      InterPro; IPR003006; IG_MHC.
DR      Pfam; PF00047; Ig; 4.
DR      SMART; SM00406; IGV; 1.
DR      PROSITE; PS00835; IG_LIKE; 4.
DR      PROSITE; PS00290; IG_MHC; 2.
KW      Hypothetical protein.
SQ      SEQUENCE 486 AA; 52662 MW; 4FEF835125DA870B CRC64;

Query Match          46.8%; Score 324; DB 4; Length 573;
Best Local Similarity 52.9%; Pred. No. 3.1e-22;
Matches 73; Conservative 15; Mismatches 38; Indels 12; Gaps 4;

QY      1 VQLQDSGGGLVQAGGSLRLSCAASGSIQDHTMGWYRQTPGKQDVVATI-QSGSTNY 59
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      20 EVQLVESGGGLVQPGKSLRLSCAASGFTFDYAMTHWVRQAPGKGLWVSGISWNSGSIGY 79
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      60 ADSVKGRTISRDNTLNTVYLQMDLKPEDTGVYVNADVPRPYTSRY-----LEIWGQ 113
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      80 ADSVKGRTISRDNAKSLYLQMSLSRAEDTALYY-----CAKGGSGSYGYGGNDVWGQ 135
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      114 GTLVTFLEP-KTPKQP 130
Db      136 GTTVTSVSSAPTKAPDVFP 153

RESULT 14
Q9Y509      PRELIMINARY;      PRT;      147 AA.
ID      Q9Y509
AC      Q9Y509
DT      01-NOV-1999 (TrEMBLrel. 12, Created)
DT      01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT      01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE      WH3 protein (Fragment).
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX      NCBI_TaxID=9606;
RN      [1]
RP      SEQUENCE FROM N.A.
RL      MEDLINE=96071149; PubMed=7475288;
RA      Cao J., Vescio R.A., Rettig M.B., Hong C.H., Kim A., Lee J.C.,
RA      Lichtenstein A.K., Berenson J.R.;
RT      "A CD10-positive subset of malignant cells is identified in multiple
RT      myeloma using PCR with patient-specific immunoglobulin gene primers."
RL      Leukemia 9:1948-1953 (1995).
DR      EMBL; S80860; AAD14339.1; -.
DR
```

```
DR HSSP; P01772; 2FB4.
DR GO; GO:0005887; C:integral to plasma membrane; NAS.
DR GO; GO:0016489; F:immunoglobulin receptor activity; NAS.
DR GO; GO:0016066; P:cellular defense response (sensu vertebrata); NAS.
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003596; IG_v.
DR Pfam; PF00047; IG; 1.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS50835; IG LIKE; 1.
FT NON_TER 1
SQ SEQUENCE 147 AA; 15768 MW; 8489FCAA7BC925C CRC64;

Query Match 46.2%; Score 319.5; DB 4; Length 147;
Best Local Similarity 57.3%; Pred. No. 1.4e-22;
Matches 71; Conservative 10; Mismatches 38; Indels 5; Gaps 2;

QY 1 QVQLDSGGGLVQAGSLRLSCAASGSIQDIHTMGWYRQTPCKQEDVVATIQSGGSTN-Y 59
Db 1 QVHVESGGGVQPGKSLRLSCAASGFTFTSYGMSWVRQAPGKGLDWWALLSYDGSTQYY 60

QY 60 ADSVKGRFTISRDNTLTNTVYLQNDLKPEDTCGVYYNADVRPRTSRY-LEIWGQGT 115
Db 61 AGSVKGRFTISRDNSKNTLYLQMTSLRVEDTAVYYCAKDNVFDSDVGYVYAGIDYWGQGT 120

QY 116 LVTV 119
Db 121 LVTV 124

RESULT 15
Q9UL88 PRELIMINARY; PRT; 131 AA.
AC Q9UL88;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Myosin-reactive immunoglobulin heavy chain variable region
DE (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98277139; PubMed=9614934;
RA Wu X., Liu B., Van der Merwe P.L., Kalis N.N., Berney S.M.,
RA Young D.C.;
RT "Myosin-reactive autoantibodies in rheumatic carditis and normal
RT fetus.";
RL Clin. Immunol. Immunopathol. 87:184-192(1998).
DR EMBL; AF035026; AAD56262.1; -.
DR PIR; S21205; S21205.
DR HSSP; P01810; 2FBJ.
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003596; IG_v.
DR Pfam; PF00047; IG; 1.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS50835; IG LIKE; 1.
FT NON_TER 1
FT NON_TER 131
SQ SEQUENCE 131 AA; 14142 MW; 96E7D668E375DEA0 CRC64;

Query Match 45.8%; Score 317; DB 4; Length 131;
Best Local Similarity 55.0%; Pred. No. 2.1e-22;
Matches 72; Conservative 12; Mismatches 33; Indels 14; Gaps 3;

QY 1 QVQLDSGGGLVQAGSLRLSCAASGSIQDIHTMGWYRQTPCKQEDVVATIQS---GGST 57
Db 1 EVQLVESGGGLVQPGKSLRLSCAASGFTFTSKAWMSWVRQAPGKGLWVGRISKTKTDGGTT 60

QY 58 NYADSVKGRFTISRDNTLTNTVYLQNDLKPEDTCGVYYNADV-----RPYRTSRYL 108
Db 61 DYAAPVKGRFTISRDSSKNTLYLRNLSLKTEDTAVYYCTTGITMIIVITTSKETS--F 118

Search completed: October 5, 2004, 08:13:49
Job time : 69.8117 secs

QY 109 EIWGGQGLTVTV 119
Db 119 EYWGGQGLTVTV 129
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2004, 07:04:10 ; Search time 98.9043 Seconds  
(without alignments)  
371.381 Million cell updates/sec

Title: US-09-805-290A-26

Perfect score: 692

Sequence: 1 QVQLQDSGGGLVQAGGSLRL.....WGQGLTVTVFLPEKTPKQPQ 130

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 segs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_29Jan04:\*

- 1: Geneseq1980s:\*
- 2: Geneseq1990s:\*
- 3: Geneseq2000s:\*
- 4: Geneseq2001s:\*
- 5: Geneseq2002s:\*
- 6: Geneseq2003as:\*
- 7: Geneseq2003bs:\*
- 8: Geneseq2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	692	100.0	130	4 AAE10559	Aae10559 HPL inhib
2	641	92.6	130	4 AAE10552	Aae10552 HPL inhib
3	617	89.2	130	4 AAE10554	Aae10554 HPL inhib
4	505.5	73.0	131	4 AAE10555	Aae10555 HPL inhib
5	502	72.5	130	4 AAE10557	Aae10557 HPL inhib
6	500	72.3	130	4 AAE10556	Aae10556 HPL inhib
7	472.5	68.3	129	4 AAE10553	Aae10553 HPL inhib
8	470	67.9	124	4 AAE10563	Aae10563 HGL inhib
9	470	67.9	152	4 AAE05283	Aae05283 Anti-pota
10	469.5	67.8	131	5 ABG30620	Abg30620 Immunoglo
11	459	66.3	124	4 AAE10566	Aae10566 HGL inhib
12	446.5	64.5	129	4 AAE10558	Aae10558 HPL inhib
13	442.5	63.9	149	4 AAE05288	Aae05288 Anti-pota
14	442	63.9	130	4 AAE10567	Aae10567 HGL inhib
15	441.5	63.8	129	4 AAE10561	Aae10561 HGL inhib
16	440	63.6	128	4 AAE10565	Aae10565 HGL inhib
17	437.5	63.2	129	4 AAE10564	Aae10564 HGL inhib
18	427.5	61.8	129	4 AAE10551	Aae10551 HPL inhib
19	422.5	61.1	118	2 AAY39819	Aay39819 Llama ant
20	422	61.0	190	7 ABR62883	Abr62883 Llama ant
21	421	60.8	190	7 ABR62882	Abr62882 Llama ant
22	419.5	60.6	377	4 AAE05286	Aae05286 Llama HCV
23	419.5	60.6	383	4 AAE05287	Aae05287 Llama HCV
24	418	60.4	190	7 ABR62881	Abr62881 Llama ant
25	417	60.3	119	4 AAB67778	Aab67778 Amino aci

26	417	60.3	119	4 AAB67777	Aab67777 Amino aci
27	417	60.3	190	7 ABR62879	Abr62879 Llama ant
28	416.5	60.2	204	2 AAY41167	Aay41167 Llama Vhh
29	416	60.1	190	7 ABR62880	Abr62880 Llama ant
30	415	60.0	198	7 ABR62878	Abr62878 Llama ant
31	413.5	59.8	211	2 AAY41172	Aay41172 Llama Vhh
32	412.5	59.6	124	2 AAY28615	Aay28615 Llama ant
33	412	59.5	134	5 ABG30618	Abg30618 Immunoglo
34	410.5	59.3	225	2 AAY41165	Aay41165 Llama Vhh
35	407	58.8	119	4 AAB67779	Aab67779 Amino aci
36	407	58.8	124	4 AAE10562	Aae10562 HGL inhib
37	406.5	58.7	153	4 AAE05282	Aae05282 Anti-pota
38	402.5	58.2	133	5 ABG30972	Abg30972 Immunoglo
39	400	57.8	130	4 AAE10560	Aae10560 HGL inhib
40	399.5	57.7	127	5 ABG30617	Abg30617 Mouse cro
41	399	57.7	194	7 ABR62885	Abr62885 Llama ant
42	398.5	57.6	205	2 AAY41173	Aay41173 Llama Vhh
43	396	57.2	111	2 AAY39816	Aay39816 Llama ant
44	392	56.6	125	2 AAY39817	Aay39817 Llama ant
45	391	56.5	206	2 AAY41169	Aay41169 Llama Vhh

#### ALIGNMENTS

RESULT 1  
AAE10559  
ID AAE10559 standard; peptide; 130 AA.  
XX  
AC AAE10559;  
XX  
DT 10-DEC-2001 (first entry)  
XX  
DE HPL inhibiting VHH fragment, HPL #30 from llama species.  
XX  
KW Llama antibody; camelid; anorectic; heavy chain variable domain; VHH;  
KW human dietary enzyme inhibitor; medicament; human pancreatic lipase; HPL;  
XX food; human gastric lipase; HGL; cosmetic control; body weight.  
XX  
OS Lama sp.

Key Location/Qualifiers  
FH Region 31..35  
FT /label= CDR1  
FT /note= "Complementarity determining region 1"  
FT Region 49..64  
FT /label= CDR2  
FT /note= "Complementarity determining region 2"  
FT Region 98..110  
FT /label= CDR3  
FT /note= "Complementarity determining region 3"

EP1134231-Al.

19-SEP-2001.

20-FEB-2001; 2001EP-00200703.

14-MAR-2000; 2000EP-00200930.

(UNIL) UNILEVER NV.

(UNIL) UNILEVER PLC.

Bezener S, Van De Burg M, De Haard JJW, Tarsilus E;

WPI; 2001-572718/65.

New antibody or its fragments for inhibiting human dietary enzymes,  
useful for cosmetic control of body weight of human beings, comprises  
heavy chain variable domain derived from immunoglobulin naturally devoid  
of light chains.

Example 2; Page 10; 37pp; English.

XX The patent discloses antibodies or their fragments comprising a heavy  
 CC chain variable domain (VHH) derived from an immunoglobulin naturally  
 CC devoid of light chains specific for inhibiting human dietary enzymes. The  
 CC antibodies of the invention are useful for the preparation of medicaments  
 CC or food for inhibiting the activity of one or more human dietary enzymes  
 CC especially human pancreatic lipase (HPL) or human gastric lipase (HGL)  
 CC which are useful for the cosmetic control of body weight of human beings.  
 CC The present peptide sequence is HPL inhibiting VHH fragment, HPL #30 from  
 CC llama (camelid) species  
 XX Sequence 130 AA;  
 SQ  
 Query Match 100.0%; Score 692; DB 4; Length 130;  
 Best Local Similarity 100.0%; Pred. No. 8.3e-57; Mismatches 0; Indels 0; Gaps 0;  
 Matches 130; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 QVQLQDSGGGLVQAGGSLRLSCAASGSIIGDIHTMGWYRQTPGKQRDVVVATIQSGGSTNYA 60  
 DB 1 QVQLQDSGGGLVQAGGSLRLSCAASGSIIGDIHTMGWYRQTPGKQRDVVVATIQSGGSTNYA 60  
 QY 61 DSVKGRFTISRDTLTNTVYLQNDLKPEDTGVVYNADVRPRTSRYLEIWGQGLTVTF 120  
 DB 61 DSVKGRFTISRDTLTNTVYLQNDLKPEDTGVVYNADVRPRTSRYLEIWGQGLTVTF 120  
 QY 121 LEPKTPKQP 130  
 DB 121 LEPKTPKQP 130  
 RESULT 2  
 AAEE10552  
 ID AAEE10552 standard; peptide; 130 AA.  
 AC AAEE10552;  
 XX  
 DT 10-DEC-2001 (first entry)  
 XX  
 DE HPL inhibiting VHH fragment, HPL #12 from llama species.  
 XX  
 KW Llama antibody; camelid; anorectic; heavy chain variable domain; VHH;  
 KW human dietary enzyme inhibitor; medicament; human pancreatic lipase; HPL;  
 KW food; human gastric lipase; HGL; cosmetic control; body weight.  
 XX  
 OS Lama sp.  
 XX  
 FH Key Location/Qualifiers  
 FT Region 31..35  
 FT /label= CDR1  
 FT /note= "Complementarity determining region 1"  
 FT Region 50..64  
 FT /label= CDR2  
 FT /note= "Complementarity determining region 2"  
 FT Region 98..110  
 FT /label= CDR3  
 FT /note= "Complementarity determining region 3"  
 XX  
 EP1134231-A1.  
 XX  
 XX 19-SEP-2001.  
 XX  
 XX 20-FEB-2001; 2001EP-00200703.  
 XX  
 XX 14-MAR-2000; 2000EP-00200930.  
 XX  
 XX (UNIL ) UNILEVER NV.  
 PA (UNIL ) UNILEVER PLC.  
 XX  
 XX Bezemer S, Van De Burg M, De Haard JJW, Tarsilus E;  
 XX WPI; 2001-572718/65.  
 XX  
 XX New antibody or its fragments for inhibiting human dietary enzymes.

PT useful for cosmetic control of body weight of human beings, comprises  
 PT heavy chain variable domain derived from immunoglobulin naturally devoid  
 PT of light chains.  
 XX  
 PS Example 2; Page 9; 37pp; English.  
 XX  
 CC The patent discloses antibodies or their fragments comprising a heavy  
 CC chain variable domain (VHH) derived from an immunoglobulin naturally  
 CC devoid of light chains specific for inhibiting human dietary enzymes. The  
 CC antibodies of the invention are useful for the preparation of medicaments  
 CC or food for inhibiting the activity of one or more human dietary enzymes  
 CC especially human pancreatic lipase (HPL) or human gastric lipase (HGL)  
 CC which are useful for the cosmetic control of body weight of human beings.  
 CC The present peptide sequence is HPL inhibiting VHH fragment, HPL #12 from  
 CC llama (camelid) species  
 XX Sequence 130 AA;  
 SQ  
 Query Match 92.6%; Score 641; DB 4; Length 130;  
 Best Local Similarity 93.1%; Pred. No. 4.8e-52;  
 Matches 121; Conservative 3; Mismatches 6; Indels 0; Gaps 0;  
 QY 1 QVQLQDSGGGLVQAGGSLRLSCAASGSIIGDIHTMGWYRQTPGKQRDVVVATIQSGGSTNYA 60  
 DB 1 QVQLQDSGGGLVQAGGSLRLSCAASGSIIGDIHTMGWYRQTPGKQRDVVVATIQSGGSTNYA 60  
 QY 61 DSVKGRFTISRDTLTNTVYLQNDLKPEDTGVVYNADVRPRTSRYLEIWGQGLTVTF 120  
 DB 61 DSVKGRFTISRDTLTNTVYLQNDLKPEDTGVVYNADVRPRTSRYLEIWGQGLTVTF 120  
 QY 121 LEPKTPKQP 130  
 DB 121 LEPKTPKQP 130  
 RESULT 3  
 AAEE10554  
 ID AAEE10554 standard; peptide; 130 AA.  
 AC AAEE10554;  
 XX  
 DT 10-DEC-2001 (first entry)  
 XX  
 DE HPL inhibiting VHH fragment, HPL #14 from llama species.  
 XX  
 KW Llama antibody; camelid; anorectic; heavy chain variable domain; VHH;  
 KW human dietary enzyme inhibitor; medicament; human pancreatic lipase; HPL;  
 KW food; human gastric lipase; HGL; cosmetic control; body weight.  
 XX  
 OS Lama sp.  
 XX  
 FH Key Location/Qualifiers  
 FT Region 31..35  
 FT /label= CDR1  
 FT /note= "Complementarity determining region 1"  
 FT Region 50..64  
 FT /label= CDR2  
 FT /note= "Complementarity determining region 2"  
 FT Region 98..110  
 FT /label= CDR3  
 FT /note= "Complementarity determining region 3"  
 XX  
 EP1134231-A1.  
 XX  
 XX 19-SEP-2001.  
 XX  
 XX 20-FEB-2001; 2001EP-00200703.  
 XX  
 XX 14-MAR-2000; 2000EP-00200930.  
 XX  
 XX (UNIL ) UNILEVER NV.  
 PA (UNIL ) UNILEVER PLC.  
 XX

```

PI Bezemer S, Van De Burg M, De Haard JJW, Tareilus E;
XX WPI; 2001-572718/65.
XX
XX New antibody or its fragments for inhibiting human dietary enzymes,
PT useful for cosmetic control of body weight of human beings, comprises
PT heavy chain variable domain derived from immunoglobulin naturally devoid
PT of light chains.
XX
XX Example 2; Page 9; 37pp; English.
XX
XX The patent discloses antibodies or their fragments comprising a heavy
CC chain variable domain (VHH) derived from an immunoglobulin naturally
CC devoid of light chains specific for inhibiting human dietary enzymes. The
CC antibodies of the invention are useful for the preparation of medicaments
CC or food for inhibiting the activity of one or more human dietary enzymes
CC especially human pancreatic lipase (HPL) or human gastric lipase (HGL)
CC which are useful for the cosmetic control of body weight of human beings.
CC The present peptide sequence is HPL inhibiting VHH fragment, HPL #14 from
CC llama (camelid) species
XX
XX Sequence 130 AA;
SQ
Query Match 89.2%; Score 617; DB 4; Length 130;
Best Local Similarity 90.0%; Pred. No. 8.3e-50;
Matches 117; Conservative 4; Mismatches 9; Indels 0; Gaps 0;
QY 1 QVQLDSGGGLVQAGGSLRSLSCAASGSGIDHTMGWYRQTPGKQRDVATIQSGGSTNYA 60
Db 1 QVQLDSGGGLVQAGGSLRSLSCAASGSGISHTMGWYRQTPGKQRDVATIQSGGSTNYA 60
QY 61 DSVKGRFTISRDNLTNTVYLQMNLDKPEDTGVYVYNNADVRPYRTSRYLLEWGGTTLTVV 120
Db 61 DSVKGRFTISRDNLTNTVYLQMNLDKPEDTGVYVYNNADVRPYRTSRYLLEWGGTTLTVV 120
QY 121 LEPKTPKQP 130
Db 121 SEPKTPKQP 130
RESULT 4
AAE10555
ID AAE10555 standard; peptide; 131 AA.
XX
XX AAE10555;
AC
XX
XX 10-DEC-2001 (first entry)
DT
XX
XX HPL inhibiting VHH fragment, HPL #15 from llama species.
DE
XX
XX Llama antibody; camelid; anorectic; heavy chain variable domain; VHH;
XX human dietary enzyme inhibitor; medicament; human pancreatic lipase; HPL;
XX food; human gastric lipase; HGL; cosmetic control; body weight.
XX
XX Llama sp.
OS
XX
XX Location/Qualifiers
FH Key 31..35
FT Region /label= CDR1
FT /note= "Complementarity determining region 1"
FT Region 50..64
FT /label= CDR2
FT /label= CDR2
FT Region 98..111
FT /label= CDR3
FT /note= "Complementarity determining region 3"
XX
XX EP1134231-A1.
XX
XX 19-SEP-2001.
PD
XX
XX 20-FEB-2001; 2001EP-00200703.
XX
XX

```

```

PR 14-MAR-2000; 2000EP-00200930.
XX
XX (UNIL ) UNILEVER NV.
PA (UNIL ) UNILEVER PLC.
XX
XX Bezemer S, Van De Burg M, De Haard JJW, Tareilus E;
XX WPI; 2001-572718/65.
XX
XX New antibody or its fragments for inhibiting human dietary enzymes,
PT useful for cosmetic control of body weight of human beings, comprises
PT heavy chain variable domain derived from immunoglobulin naturally devoid
PT of light chains.
XX
XX Example 2; Page 9; 37pp; English.
XX
XX The patent discloses antibodies or their fragments comprising a heavy
CC chain variable domain (VHH) derived from an immunoglobulin naturally
CC devoid of light chains specific for inhibiting human dietary enzymes. The
CC antibodies of the invention are useful for the preparation of medicaments
CC or food for inhibiting the activity of one or more human dietary enzymes
CC especially human pancreatic lipase (HPL) or human gastric lipase (HGL)
CC which are useful for the cosmetic control of body weight of human beings.
CC The present peptide sequence is HPL inhibiting VHH fragment, HPL #15 from
CC llama (camelid) species
XX
XX Sequence 131 AA;
SQ
Query Match 73.0%; Score 505.5; DB 4; Length 131;
Best Local Similarity 74.8%; Pred. No. 2.2e-39;
Matches 98; Conservative 11; Mismatches 21; Indels 1; Gaps 1;
QY 1 QVQLDSGGGLVQAGGSLRSLSCAASGSGIDHTMGWYRQTPGKQRDVATIQSGGSTNYA 60
Db 1 QVQLDSGGGLVQAGGSLRSLSCAASGSGISSINVMGWFPQAFKQRELVASITSGGSTNYA 60
QY 61 DSVKGRFTISRDNLTNTVYLQMNLDKPEDTGVYVYNNADVRPYRTSRYLLEWGGTTLTVV 119
Db 61 DSLKGRFTISRDNKNAVYLVQMNLDKPEDTAVYCNHAITPAGSSNYVYVYVYVYVYVYVYVYV 120
QY 120 FLEPKTPKQP 130
Db 121 SSEPKTPKQP 131
RESULT 5
AAE10557
ID AAE10557 standard; peptide; 130 AA.
XX
XX AAE10557;
AC
XX
XX 10-DEC-2001 (first entry)
DT
XX
XX HPL inhibiting VHH fragment, HPL #19 from llama species.
DE
XX
XX Llama antibody; camelid; anorectic; heavy chain variable domain; VHH;
XX human dietary enzyme inhibitor; medicament; human pancreatic lipase; HPL;
XX food; human gastric lipase; HGL; cosmetic control; body weight.
XX
XX Llama sp.
OS
XX
XX Location/Qualifiers
FH Key 31..35
FT Region /label= CDR1
FT /note= "Complementarity determining region 1"
FT Region 50..64
FT /label= CDR2
FT /note= "Complementarity determining region 2"
FT Region 98..110
FT /label= CDR3
FT /note= "Complementarity determining region 3"
XX
XX EP1134231-A1.
XX

```

```

XX 19-SEP-2001.
XX 20-FEB-2001; 2001EP-00200703.
XX 14-MAR-2000; 2000EP-00200930.
XX (UNIL ) UNILEVER NV.
XX (UNIL ) UNILEVER PLC.
XX Bezemer S, Van De Burg M, De Haard JJW, Tareilus E;
XX WPI; 2001-572718/65.
XX New antibody or its fragments for inhibiting human dietary enzymes,
XX useful for cosmetic control of body weight of human beings, comprises
XX heavy chain variable domain derived from immunoglobulin naturally devoid
XX of light chains.
XX Example 2; Page 10; 37pp; English.
XX The patent discloses antibodies or their fragments comprising a heavy
XX chain variable domain (VHH) derived from an immunoglobulin naturally
XX devoid of light chains specific for inhibiting human dietary enzymes. The
XX antibodies of the invention are useful for the preparation of medicaments
XX or food for inhibiting the activity of one or more human dietary enzymes
XX especially human pancreatic lipase (HPL) or human gastric lipase (HGL)
XX which are useful for the cosmetic control of body weight of human beings.
XX The present peptide sequence is HPL inhibiting VHH fragment, HPL #19 from
XX llama (camelid) species
XX Sequence 130 AA;
Query Match 72.5%; Score 502; DB 4; Length 130;
Best Local Similarity 74.6%; Pred. No. 4.5e-39;
Matches 97; Conservative 8; Mismatches 25; Indels 0; Gaps 0;
QY 1 QVQLDSGGGLVQAGGSLRLSCAASGSIQDIHTMGWYRQTPGKQRDVVATISGGSTNYA 60
DB 1 QVQLDSGGGLVQAGGSLRLSCAASGSIQDIHTMGWYRQTPGKQRDVVATISGGSTNYA 60
QY 61 DSVKGRFTISRNTLTNTVYLOMNDLKPEDTGVYVYNNADVPYRYSRYLEIWGGSTLVTVF 120
DB 61 DSVKGRFTISRNTLTNTVYLOMNDLKPEDTGVYVYNNADVPYRYSRYLEIWGGSTLVTVF 120
QY 121 LEPTKPKQP 130
DB 121 SEPTKPKQP 130
RESULT 6
AAE10556
ID AAE10556 standard; peptide; 130 AA.
XX AAE10556;
XX 10-DEC-2001 (first entry)
XX HPL inhibiting VHH fragment, HPL #18 from llama species.
XX Llama antibody; camelid; anorectic; heavy chain variable domain; VHH;
XX human dietary enzyme inhibitor; medicament; human pancreatic lipase; HPL;
XX food; human gastric lipase; HGL; cosmetic control; body weight.
XX Lama sp.
XX Location/Qualifiers
XX Key 31..35
XX Region /label= CDR1
XX 50..64
XX Region /label= CDR2
XX /note= "Complementarity determining region 1"
XX /note= "Complementarity determining region 2"

```

```

FT Region 98..110
FT /label= CDR3
FT /note= "Complementarity determining region 3"
XX EPI134231-A1.
XX 19-SEP-2001.
XX 20-FEB-2001; 2001EP-00200703.
XX 14-MAR-2000; 2000EP-00200930.
XX (UNIL ) UNILEVER NV.
XX (UNIL ) UNILEVER PLC.
XX Bezemer S, Van De Burg M, De Haard JJW, Tareilus E;
XX WPI; 2001-572718/65.
XX New antibody or its fragments for inhibiting human dietary enzymes,
XX useful for cosmetic control of body weight of human beings, comprises
XX heavy chain variable domain derived from immunoglobulin naturally devoid
XX of light chains.
XX Example 2; Page 10; 37pp; English.
XX The patent discloses antibodies or their fragments comprising a heavy
XX chain variable domain (VHH) derived from an immunoglobulin naturally
XX devoid of light chains specific for inhibiting human dietary enzymes. The
XX antibodies of the invention are useful for the preparation of medicaments
XX or food for inhibiting the activity of one or more human dietary enzymes
XX especially human pancreatic lipase (HPL) or human gastric lipase (HGL)
XX which are useful for the cosmetic control of body weight of human beings.
XX The present peptide sequence is HPL inhibiting VHH fragment, HPL #18 from
XX llama (camelid) species
XX Sequence 130 AA;
Query Match 72.3%; Score 500; DB 4; Length 130;
Best Local Similarity 75.4%; Pred. No. 7e-39;
Matches 99; Conservative 8; Mismatches 24; Indels 0; Gaps 0;
QY 1 QVQLDSGGGLVQAGGSLRLSCAASGSIQDIHTMGWYRQTPGKQRDVVATISGGSTNYA 60
DB 1 QVQLDSGGGLVQAGGSLRLSCAASGSIQDIHTMGWYRQTPGKQRDVVATISGGSTNYA 60
QY 61 DSVKGRFTISRNTLTNTVYLOMNDLKPEDTGVYVYNNADVPYRYSRYLEIWGGSTLVTVF 120
DB 61 DSVKGRFTISRNTLTNTVYLOMNDLKPEDTGVYVYNNADVPYRYSRYLEIWGGSTLVTVF 120
QY 121 LEPTKPKQP 130
DB 121 SEPTKPKQP 130
RESULT 7
AAE10553
ID AAE10553 standard; peptide; 129 AA.
XX AAE10553;
XX 10-DEC-2001 (first entry)
XX HPL inhibiting VHH fragment, HPL #13 from llama species.
XX Llama antibody; camelid; anorectic; heavy chain variable domain; VHH;
XX human dietary enzyme inhibitor; medicament; human pancreatic lipase; HPL;
XX food; human gastric lipase; HGL; cosmetic control; body weight.
XX Lama sp.
XX Location/Qualifiers
XX Key 31..35
XX Region

```

FT /label= CDR1  
 FT /note= "Complementarity determining region 1"  
 FT 50..64  
 FT Region  
 FT /label= CDR2  
 FT /note= "Complementarity determining region 2"  
 FT 98..109  
 FT Region  
 FT /label= CDR3  
 FT /note= "Complementarity determining region 3"  
 FT  
 XX EP1134231-A1.  
 XX  
 XX  
 PD 19-SEP-2001.  
 XX  
 XX 20-FEB-2001; 2001EP-00200703.  
 XX  
 XX 14-MAR-2000; 2000EP-00200930.  
 XX  
 XX (UNIL ) UNILEVER NV.  
 XX (UNIL ) UNILEVER PLC.  
 XX  
 XX Bezemer S, Van De Burg M, De Haard JJW, Tareilus E;  
 XX WPI; 2001-572718/65.  
 XX  
 XX New antibody or its fragments for inhibiting human dietary enzymes,  
 XX useful for cosmetic control of body weight of human beings, comprises  
 XX heavy chain variable domain derived from immunoglobulin naturally devoid  
 XX of light chains.

Example 2; Page 9; 37pp; English.

The patent discloses antibodies or their fragments comprising a heavy chain variable domain (VHH) derived from an immunoglobulin naturally devoid of light chains specific for inhibiting human dietary enzymes. The antibodies of the invention are useful for the preparation of medicaments or food for inhibiting the activity of one or more human dietary enzymes especially human pancreatic lipase (HPL) or human gastric lipase (HGL) which are useful for the cosmetic control of body weight of human beings. The present peptide sequence is HPL inhibiting VHH fragment, HPL #13 from llama (camelid) species

Sequence 129 AA;

Query Match 68.3%; Score 472.5; DB 4; Length 129;  
 Best Local Similarity 73.1%; Pred. No. 2.6e-36;  
 Matches 95; Conservative 9; Mismatches 25; Indels 1; Gaps 1;  
 QY 1 QVQLDSGGGLVQAGGSLRLSCAASGSIIGDIHTMGWYRQTGKQRDVVAITQSGGSTNYA 60  
 Db 1 QVQLDSGGGLVQAGGSLRLSCAASGSIIGDIHTMGWYRQTGKQRELVGRITAGGSTNYA 60  
 QY 61 DSVKGRFTISRDNLTNTVYLQNDLKPEDTGVYYNADVRPRTSRYLEIWGGQTLTVTF 120  
 Db 61 DSAKGRFTISKDAKNTVYLQNSLKPEDTAVYICNALITFWDKS-VNDYWGQGTQVTVS 119  
 QY 121 LEPTKPKQP 130  
 Db 120 SEPTKPKQP 129

RESULT 8  
 AAE10563  
 ID AAE10563 standard; peptide; 124 AA.  
 XX  
 AC AAE10563;  
 XX  
 XX 10-DEC-2001 (first entry)  
 XX  
 XX HGL inhibiting VHH fragment, HGL #9 from llama species.

Llama antibody; camelid; anorectic; heavy chain variable domain; VHH;  
 KW human dietary enzyme inhibitor; medicament; human pancreatic lipase; HPL;  
 KW food; human gastric lipase; HGL; cosmetic control; body weight.

XX Lama sp.  
 XX OS  
 XX Key  
 XX Location/Qualifiers  
 FT 31..35  
 FT Region  
 FT /label= CDR1  
 FT /note= "Complementarity determining region 1"  
 FT 50..64  
 FT Region  
 FT /label= CDR2  
 FT /note= "Complementarity determining region 2"  
 FT 98..104  
 FT Region  
 FT /label= CDR3  
 FT /note= "Complementarity determining region 3"  
 FT  
 XX EP1134231-A1.  
 XX  
 XX 19-SEP-2001.  
 XX  
 XX 20-FEB-2001; 2001EP-00200703.  
 XX  
 XX 14-MAR-2000; 2000EP-00200930.  
 XX  
 XX (UNIL ) UNILEVER NV.  
 XX (UNIL ) UNILEVER PLC.  
 XX  
 XX Bezemer S, Van De Burg M, De Haard JJW, Tareilus E;  
 XX WPI; 2001-572718/65.  
 XX  
 XX New antibody or its fragments for inhibiting human dietary enzymes,  
 XX useful for cosmetic control of body weight of human beings, comprises  
 XX heavy chain variable domain derived from immunoglobulin naturally devoid  
 XX of light chains.

Example 4; Page 13; 37pp; English.

The patent discloses antibodies or their fragments comprising a heavy chain variable domain (VHH) derived from an immunoglobulin naturally devoid of light chains specific for inhibiting human dietary enzymes. The antibodies of the invention are useful for the preparation of medicaments or food for inhibiting the activity of one or more human dietary enzymes especially human pancreatic lipase (HPL) or human gastric lipase (HGL) which are useful for the cosmetic control of body weight of human beings. The present peptide sequence is HGL inhibiting VHH fragment, HGL #9 from llama (camelid) species

Sequence 124 AA;

Query Match 67.9%; Score 470; DB 4; Length 124;  
 Best Local Similarity 71.5%; Pred. No. 4.2e-36;  
 Matches 93; Conservative 6; Mismatches 25; Indels 6; Gaps 1;  
 QY 1 QVQLDSGGGLVQAGGSLRLSCAASGSIIGDIHTMGWYRQTGKQRDVVAITQSGGSTNYA 60  
 Db 1 QVQLDSGGGLVQAGGSLRLSCAASGSIIGSLVMSWYRQAPGKQREPVAALMGSGSTTYA 60  
 QY 61 DSVKGRFTISRDNLTNTVYLQNDLKPEDTGVYYNADVRPRTSRYLEIWGGQTLTVTF 120  
 Db 61 DSVKGRFTISRDNLTNTVYLQNSLKPEDTGVYYCAG-----TGAEGHWGGGTQVTVS 114  
 QY 121 LEPTKPKQP 130  
 Db 115 SEPTKPKQP 124

RESULT 9  
 AAE05283  
 ID AAE05283 standard; protein; 152 AA.  
 XX  
 AC AAE05283;  
 XX  
 XX 18-SEP-2001 (first entry)  
 XX

DE Anti-potato SBEII (Clone46) VH region attached with myc and his6 tag.  
 XX  
 KW Potato; heavy chain immunoglobulin; pathogen resistance;  
 KW metabolism modulator; passive immunisation; heavy chain variable domain;  
 KW VH; anti-potato SBEII; starch branching enzyme; SBE A.  
 XX  
 OS Solanum tuberosum.  
 OS Unidentified.  
 OS Chimeric.  
 XX  
 PN EP1118669-A2.  
 XX  
 XX 25-JUL-2001.  
 XX  
 XX 08-DEC-2000; 2000EP-00310997.  
 XX  
 XX 17-DEC-1999; 99EP-00310188.  
 XX  
 XX (UNIL ) UNILEVER PLC.  
 XX (UNIL ) UNILEVER NV.  
 XX  
 XX Frenken LGJ, Van Der Logt CPE, Jobling SA, Teh Y;  
 PI  
 XX WPI; 2001-427157/46.  
 DR N-PSDB; AAD10054.  
 DR  
 XX  
 XX Modifying a plant to produce an antibody useful for increasing pathogen  
 PT resistance or to modulate metabolism comprises introducing a DNA sequence  
 PT encoding a heavy chain immunoglobulin linked to a peptide that targets a  
 PT cellular compartment.  
 XX  
 XX Example 1; Fig 12; 81pp; English.  
 XX  
 XX The present invention relates to a method for modifying a plant to  
 CC produce an antibody or an active fragment or derivative, or a protein  
 CC functional equivalent, in a cellular compartment comprising introducing a  
 CC DNA sequence encoding a heavy chain immunoglobulin, where the DNA is  
 CC linked to promoters and provided with an additional sequence encoding a  
 CC peptide capable of targeting heavy chain immunoglobulin to a cellular  
 CC compartment. The method is used for producing a heavy chain  
 CC immunoglobulin or an active fragment or derivative, or a protein that is  
 CC functionally equivalent for increasing the pathogen resistance in a plant  
 CC or to modulate metabolism in a plant. Under some circumstances it may be  
 CC desirable to retain the antibody product with the plant rather than  
 CC extracting and isolating the product. In particular, edible selected  
 CC antigens may be used in a method of passively immunising an animal,  
 CC preferably human, against the antigen, e.g., pathogenic organisms. The  
 CC present sequence is an anti-potato SBEII antibody (denoted Clone46) heavy  
 CC chain variable domain (VH) attached to peptide linkers, myc and his6 tag.  
 CC The potato SBEII is a starch branching enzyme also designated SBE A

XX Sequence 152 AA;

Query Match 67.9%; Score 470; DB 4; Length 152;  
 Best Local Similarity 72.4%; Pred. No. 5,3e-36;  
 Matches 97; Conservative 5; Mismatches 20; Indels 12; Gaps 2;  
 QY 1 QVQLQDSGGELVQAGSLRLSCAAGSGTGDITMGWYRQTPGKQDVVATIOSGGSNYA 60  
 DB 1 QVQLQESGGELVQAGSLRLSCVASFSTFIAMAWYRQAPKQREVVASINSIGSYA 60  
 QY 61 DSVKGRFTISRNTLTNTVYQWMDLKPDGTGYVYNAD- ---VPRYRSTRYLETWGGQTL 116  
 DB 61 DSVKGRFTISRNTAKNTVYQWSSSKPDYAVYCAAGNLVLRPY-----WGQGTLL 112  
 QY 117 VTFVLEPKTPKQP 130  
 DB 113 VTVSSEPKTPKQP 126

RESULT 10  
 ABG30620 standard; protein; 131 AA.  
 ID ABG30620

XX AC ABG30620;  
 XX  
 XX 21-OCT-2002 (first entry)  
 XX  
 XX Immunoglobulin G specific heavy chain variable domain antibody #3.  
 XX  
 XX Heavy chain variable domain; antibody; protein array; aging; VHH;  
 KW immunoglobulin; sticky phage-antibody; micro-panning; immune library;  
 KW proteomic; mouse.  
 XX  
 XX Mus sp.  
 OS  
 XX WO200248193-A2.  
 PN  
 XX 20-JUN-2002.  
 PD  
 XX 03-DEC-2001; 2001WO-EP014471.  
 XX  
 XX 13-DEC-2000; 2000EP-00311142.  
 PR  
 XX (UNIL ) UNILEVER NV.  
 XX (UNIL ) UNILEVER PLC.  
 XX (UNIL ) HINDUSTAN LEVER LTD.  
 XX  
 XX De Haard JJW, Hermans P, Landa I, Verrips CT;  
 PI  
 XX WPI; 2002-583487/62.  
 DR  
 XX Novel protein array useful for detecting the presence of individual  
 PT proteins in sample, comprises heavy-chain variable domain antibodies or  
 PT antibody fragments obtainable from Camelidae.  
 XX  
 XX Disclosure; Fig 3; 80pp; English.  
 PS  
 XX The invention describes a protein array (I) comprising a number of heavy-  
 CC chain variable domain antibodies or antibody fragments, obtainable from  
 CC Camelidae. The method is useful for removing abundant proteins from an  
 CC extract or sample which do not provide useful information on the  
 CC condition of a cell or tissue in the extract or sample to be  
 CC investigated. (I) is useful for detecting the presence of individual  
 CC proteins in a sample, comparing the distribution of proteins in different  
 CC cell types, and identification of proteins that may be of importance in  
 CC determining the altered properties of cells in disease, aging or other  
 CC conditions. Using a heavy-chain variable domain derived from an  
 CC immunoglobulin that is naturally devoid of light chains (VHH) in (I)  
 CC provides a number of advantages, such as an improvement of  
 CC sensitivity/resolution in the order of 10-100 times, and detection of  
 CC post-translationally modified proteins. The invention also describes a  
 CC method (II) that enables the simultaneous processing of large numbers of  
 CC target antigens in a controlled way. The incorporated phage-ELISA  
 CC generates on-line information about the success or failure of a certain  
 CC panning condition. This feature combined with the microtiter plate format  
 CC allows the complete automation of the technology, based on computer-made  
 CC decisions on the values of the phage-ELISA for continuation of a limited  
 CC number of selections. In (II), due to the fact that many different  
 CC conditions can be tested, varying amounts of input-phages can be used  
 CC simultaneously in order to decrease the enrichment of sticky phage-  
 CC antibodies. Micro-panning is an effective tool for selecting both naive,  
 CC synthetic and immune libraries on large numbers of different target  
 CC molecules, enabling the generation of large panels of antibodies in short  
 CC time frames needed for the generation of arrays (proteomics). This  
 CC sequence represents an immunoglobulin (IgG) heavy chain variable domain  
 CC antibody (VHH)

Sequence 131 AA;

Query Match 67.8%; Score 469.5; DB 5; Length 131;  
 Best Local Similarity 71.0%; Pred. No. 5e-36;  
 Matches 93; Conservative 10; Mismatches 27; Indels 1; Gaps 1;

QY 1 QVQLQDSGGELVQAGSLRLSCAAGSGTGDITMGWYRQTPGKQDVVATIOSGGSNYA 60  
 ID ABG30620



```

Db      1 QVQLQESGGGLVQPGGSLRLSCAASKSIIFGFGAVGHWHRQAPGKQRELVARITYDSGTNYA 60
QY      61 DSVKGRFTISRDNLTNTVYLQMDLKPEDTGVVYMNADVPRPYSRYL-EIMQGGLTVTV 119
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      61 DSVKGRFTISRDNLTNTVYLQMDLKPEDTGVVYMNADVPRPYSRYL-EIMQGGLTVTV 120
QY      120 FLPEKTPKPOP 130
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      121 SSEPKTTPKPOP 131
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 11
AAE10566
ID      AAE10566 standard; peptide; 124 AA.
XX      AC      AAE10566;
XX      XX
XX      DT      10-DEC-2001 (first entry)
XX      DE      HGL inhibiting VHH fragment, HGL #15 from llama species.
XX      KW      Llama antibody; camelid; anorectic; heavy chain variable domain; VHH;
XX      KW      human dietary enzyme inhibitor; medicament; human pancreatic lipase; HPL;
XX      KW      food; human gastric lipase; HGL; cosmetic control; body weight.
XX      OS      Llama sp.
XX      FH      Key      Location/Qualifiers
XX      FT      Region      30..35
XX      FT      /label= CDR1
XX      FT      /note= "Complementarity determining region 1"
XX      FT      Region      50..64
XX      FT      /label= CDR2
XX      FT      /note= "Complementarity determining region 2"
XX      FT      Region      98..104
XX      FT      /label= CDR3
XX      FT      /note= "Complementarity determining region 3"
XX      EP1134231-A1.
XX      DD      19-SEP-2001.
XX      FF      20-FEB-2001; 2001EP-00200703.
XX      PR      14-MAR-2000; 2000EP-00200930.
XX      PA      (UNIL ) UNILEVER NV.
XX      PA      (UNIL ) UNILEVER PLC.
XX      PI      Bezemer S, Van De Burg M, De Haard JJW, Tareilus E;
XX      DR      WPI; 2001-572718/65.
XX      PT      New antibody or its fragments for inhibiting human dietary enzymes,
XX      PT      useful for cosmetic control of body weight of human beings, comprises
XX      PT      heavy chain variable domain derived from immunoglobulin naturally devoid
XX      PT      of light chains.
XX      PS      Example 4; Page 14; 37pp; English.
XX      CC      The patent discloses antibodies or their fragments comprising a heavy
XX      CC      chain variable domain (VHH) derived from an immunoglobulin naturally
XX      CC      devoid of light chains specific for inhibiting human dietary enzymes. The
XX      CC      antibodies of the invention are useful for the preparation of medicaments
XX      CC      or food for inhibiting the activity of one or more human dietary enzymes
XX      CC      especially human pancreatic lipase (HPL) or human gastric lipase (HGL)
XX      CC      which are useful for the cosmetic control of body weight of human beings.
XX      CC      The present peptide sequence is HGL inhibiting VHH fragment, HGL #15 from
XX      CC      llama (camelid) species
XX      SQ      Sequence 124 AA;
        Query Match      66.3%; Score 459; DB 4; Length 124;

```

```

Best Local Similarity 69.2%; Pred. No. 4.4e-35;
Matches 90; Conservative 8; Mismatches 26; Indels 6; Gaps 1;
QY      1 QVQLQESGGGLVQAGGSLRLSCAASGSIQDTHTWGWRQTGKQKRDVVATIQSGGSTNYA 60
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      1 QVQLQESGGGLVQAGGSLRLSCAASGSIQSGMYVMSWYRQAPGKEREPTAALMGSGSTTYA 60
QY      61 DSVKGRFTISRDNLTNTVYLQMDLKPEDTGVVYMNADVPRPYSRYL-EIMQGGLTVTV 120
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      61 DSVKGRFTISRDNLTNTVYLQMDLKPEDTGVVYMNADVPRPYSRYL-EIMQGGLTVTV 114
QY      121 LEPTKPKPOP 130
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      115 SEPKTTPKPOP 124
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 12
AAE10558
ID      AAE10558 standard; peptide; 129 AA.
XX      AC      AAE10558;
XX      XX
XX      DT      10-DEC-2001 (first entry)
XX      DE      HPL inhibiting VHH fragment, HPL #22 from llama species.
XX      KW      Llama antibody; camelid; anorectic; heavy chain variable domain; VHH;
XX      KW      human dietary enzyme inhibitor; medicament; human pancreatic lipase; HPL;
XX      KW      food; human gastric lipase; HGL; cosmetic control; body weight.
XX      OS      Llama sp.
XX      FH      Key      Location/Qualifiers
XX      FT      Region      31..35
XX      FT      /label= CDR1
XX      FT      /note= "Complementarity determining region 1"
XX      FT      Region      50..64
XX      FT      /label= CDR2
XX      FT      /note= "Complementarity determining region 2"
XX      FT      Region      98..109
XX      FT      /label= CDR3
XX      FT      /note= "Complementarity determining region 3"
XX      EP1134231-A1.
XX      DD      19-SEP-2001.
XX      FF      20-FEB-2001; 2001EP-00200703.
XX      PR      14-MAR-2000; 2000EP-00200930.
XX      PA      (UNIL ) UNILEVER NV.
XX      PA      (UNIL ) UNILEVER PLC.
XX      PI      Bezemer S, Van De Burg M, De Haard JJW, Tareilus E;
XX      DR      WPI; 2001-572718/65.
XX      PT      New antibody or its fragments for inhibiting human dietary enzymes,
XX      PT      useful for cosmetic control of body weight of human beings, comprises
XX      PT      heavy chain variable domain derived from immunoglobulin naturally devoid
XX      PT      of light chains.
XX      PS      Example 2; Page 10; 37pp; English.
XX      CC      The patent discloses antibodies or their fragments comprising a heavy
XX      CC      chain variable domain (VHH) derived from an immunoglobulin naturally
XX      CC      devoid of light chains specific for inhibiting human dietary enzymes. The
XX      CC      antibodies of the invention are useful for the preparation of medicaments
XX      CC      or food for inhibiting the activity of one or more human dietary enzymes
XX      CC      especially human pancreatic lipase (HPL) or human gastric lipase (HGL)
XX      CC      which are useful for the cosmetic control of body weight of human beings.
XX      CC      The present peptide sequence is HPL inhibiting VHH fragment, HPL #22 from

```

CC llama (camelid) species  
 XX Sequence 129 AA;  
 SQ Query Match 64.5%; Score 446.5; DB 4; Length 129;  
 Best Local Similarity 69.2%; Pred. No. 6.8e-34;  
 Matches 90; Conservative 12; Mismatches 27; Indels 1; Gaps 1;

QY 1 QVQLDSGGGLVQAGGSLRLSCAASGSIIDITMGWYRTPGKQDVVATIQSGSTNYA 60  
 Db 1 QVQLDSGGGLVQAGGSLRLSCAASGSIISIMTWYRQAPGKERELVARMSSDGTTSYT 60  
 QY 61 DSVKGRFTISRDNLTNTVYLNQNDLKPEDTGVYNNADVRPYRTSRYLEIMQGGLTVTF 120  
 Db 61 DSVKGRFTISRDNKNTVYLNQNDLKPEDTAVYCKALISSYDGS-WNDYGGGQTQVTVS 119

QY 121 LEPKTPKQP 130  
 Db 120 SEPKTPKQP 129

RESULT 13  
 AAE05288  
 ID AAE05288 standard; protein; 149 AA.  
 XX AAE05288;  
 AC AAE05288;  
 DT 18-SEP-2001 (first entry)  
 XX Anti-potato SBEII (Clone68) VH region attached with myc and his6 tag.  
 DE Potato; heavy chain immunoglobulin; pathogen resistance;  
 KW metabolism modulator; passive immunisation; heavy chain variable domain;  
 KW VH; anti-potato SBEII; starch branching enzyme; SBE A.  
 XX Solanum tuberosum.  
 OS Unidentified.  
 OS Chimeric.  
 XX EP1118669-A2.  
 XX 25-JUL-2001.  
 XX 08-DEC-2000; 2000EP-00310997.  
 XX 17-DEC-1999; 99EP-00310189.  
 XX (UNIL) UNILEVER PLC.  
 XX (UNIL) UNILEVER NV.  
 XX Franken LGJ, Van Der Logt CPE, Jobling SA, Teh Y;  
 PI WPI; 2001-427157/46.  
 XX N-PSDB; AAD10059.  
 XX Modifying a plant to produce an antibody useful for increasing pathogen  
 PT resistance or to modulate metabolism comprises introducing a DNA sequence  
 PT encoding a heavy chain immunoglobulin linked to a peptide that targets a  
 PT cellular compartment.  
 XX Example 13; Fig 28; 81pp; English.  
 XX The present invention relates to a method for modifying a plant to  
 CC produce an antibody or an active fragment or derivative, or a protein  
 CC functional equivalent, in a cellular compartment comprising introducing a  
 CC DNA sequence encoding a heavy chain immunoglobulin, where the DNA is  
 CC linked to promoters and provided with an additional sequence encoding a  
 CC peptide capable of targeting heavy chain immunoglobulin to a cellular  
 CC compartment. The method is used for producing a heavy chain  
 CC immunoglobulin or an active fragment or derivative, or a protein that is  
 CC functionally equivalent for increasing the pathogen resistance in a plant  
 CC or to modulate metabolism in a plant. Under some circumstances it may be  
 CC desirable to retain the antibody product with the plant rather than

CC extracting and isolating the product. In particular, edible selected  
 CC antigens may be used in a method of passively immunising an animal,  
 CC preferably human, against the antigen, e.g., pathogenic organisms. The  
 CC present sequence is an anti-potato SBEII antibody (denoted Clone68) heavy  
 CC chain variable domain (VH) attached to peptide linkers, myc and his6 tag.  
 CC The potato SBEII is a starch branching enzyme also designated SBE A  
 XX Sequence 149 AA;  
 SQ Query Match 63.9%; Score 442.5; DB 4; Length 149;  
 Best Local Similarity 69.2%; Pred. No. 1.9e-33;  
 Matches 90; Conservative 8; Mismatches 23; Indels 9; Gaps 1;

QY 1 QVQLDSGGGLVQAGGSLRLSCAASGSIIDITMGWYRTPGKQDVVATIQSGSTNYA 60  
 Db 3 QVQLDSGGGLVQAGGSLRLSCAASGSIIFRRPHNGWFRQAPGQERELVALISAGGTWYA 62  
 QY 61 DSVKGRFTISRDNLTNTVYLNQNDLKPEDTGVYNNADVRPYRTSRYLEIMQGGLTVTF 120  
 Db 63 DSVKGRFTISRDNKNTVYLNQNDLKPEDTAVYCTAGG-----SYWGGGQTQVTV 113

QY 121 LEPKTPKQP 130  
 Db 114 SEPKTPKQP 123

RESULT 14  
 AAE10567  
 ID AAE10567 standard; peptide; 130 AA.  
 XX AAE10567;  
 AC AAE10567;  
 DT 10-DEC-2001 (first entry)  
 XX HGL inhibiting VHH fragment, HGL #16 from llama species.  
 DE Llama antibody; camelid; anorectic; heavy chain variable domain; VHH;  
 KW human dietary enzyme inhibitor; medicament; human pancreatic lipase; HPL;  
 KW food; human gastric lipase; HGL; cosmetic control; body weight.  
 XX Lama sp.  
 OS Key Location/Qualifiers  
 FH Region 31..35  
 FT /label= CDR1  
 FT /note= "Complementarity determining region 1"  
 FT Region 50..64  
 FT /label= CDR2  
 FT /note= "Complementarity determining region 2"  
 FT Region 98..110  
 FT /label= CDR3  
 FT /note= "Complementarity determining region 3"  
 XX EP1134231-A1.  
 XX 19-SEP-2001.  
 XX 20-FEB-2001; 2001EP-00200703.  
 XX 14-MAR-2000; 2000EP-00200930.  
 XX (UNIL) UNILEVER NV.  
 XX (UNIL) UNILEVER PLC.  
 XX Bezemer S, Van De Burg M, De Haard JJW, Tareilus E;  
 WPI; 2001-572718/65.  
 XX New antibody or its fragments for inhibiting human dietary enzymes,  
 PT useful for cosmetic control of body weight of human beings, comprises  
 PT heavy chain variable domain derived from immunoglobulin naturally devoid  
 PT of light chains.  
 XX

PS Example 4; Page 14; 37pp; English.

XX The patent discloses antibodies or their fragments comprising a heavy

CC chain variable domain (VHH) derived from an immunoglobulin naturally

CC devoid of light chains specific for inhibiting human dietary enzymes. The

CC antibodies of the invention are useful for the preparation of medicaments

CC or food for inhibiting the activity of one or more human dietary enzymes

CC especially human pancreatic lipase (HPL) or human gastric lipase (HGL)

CC which are useful for the cosmetic control of body weight of human beings.

CC The present peptide sequence is HGL inhibiting VHH fragment, HGL #16 from

CC llama (camelid) species

XX Sequence 130 AA;

SQ

Query Match 63.9%; Score 442; DB 4; Length 130;

Best Local Similarity 70.8%; Pred. No. 1.8e-33;

Matches 92; Conservative 4; Mismatches 34; Indels 0; Gaps 0;

QY 1 QVQLDSGGGLVQAGGSLRLSCAASGSIQDIHTWGYRQTPGKQRDVAVTIQSGGSTNYA 60

Db 1 QVQLDSGGGLVQAGGSLRLSCAASGSIQDIHTWGYRQTPGKQKLVATITTYRTNYA 60

QY 61 DSVKGRFTISRDNLTNTLYLQNDLKPEDTGVYNNADVPYRTSRYLEIWGGTTLVTVF 120

Db 61 DSVKGRFTISRDNLTNTLYLQNDLKPEDTGVYNNADVPYRTSRYLEIWGGTTLVTVF 120

QY 121 LEPTKPKQP 130

Db 121 SEPTKPKQP 130

RESULT 15

AAE10561

ID AAE10561 standard; peptide; 129 AA.

XX AAE10561;

XX 10-DEC-2001 (first entry)

XX HGL inhibiting VHH fragment, HGL #4 from llama species.

XX Llama antibody; camelid; anorectic; heavy chain variable domain; VHH;

KW human dietary enzyme inhibitor; medicament; human pancreatic lipase; HPL;

KW food; human gastric lipase; HGL; cosmetic control; body weight.

XX Llama sp.

XX Key Location/Qualifiers

FT Region 31..35

FT /label= CDR1

FT /note= "Complementarity determining region 1"

FT Region 50..64

FT /label= CDR2

FT /note= "Complementarity determining region 2"

FT Region 98..109

FT /label= CDR3

FT /note= "Complementarity determining region 3"

XX EP1134231-A1.

FN

XX

PD 19-SEP-2001.

XX

PF 20-FEB-2001; 2001EP-00200703.

XX

PR 14-MAR-2000; 2000EP-00200930.

XX

PA (UNIL ) UNILEVER NV.

PA (UNIL ) UNILEVER PLC.

XX

PI Bezemer S, Van De Burg M, De Haard JJW, Tareilus E;

XX

DR WPI; 2001-572718/65.

XX

PT New antibody or its fragments for inhibiting human dietary enzymes,

PT useful for cosmetic control of body weight of human beings, comprises

PT heavy chain variable domain derived from immunoglobulin naturally devoid

PT of light chains.

XX Example 4; Page 13; 37pp; English.

XX The patent discloses antibodies or their fragments comprising a heavy

CC chain variable domain (VHH) derived from an immunoglobulin naturally

CC devoid of light chains specific for inhibiting human dietary enzymes. The

CC antibodies of the invention are useful for the preparation of medicaments

CC or food for inhibiting the activity of one or more human dietary enzymes

CC especially human pancreatic lipase (HPL) or human gastric lipase (HGL)

CC which are useful for the cosmetic control of body weight of human beings.

CC The present peptide sequence is HGL inhibiting VHH fragment, HGL #4 from

CC llama (camelid) species

XX Sequence 129 AA;

SQ

Query Match 63.8%; Score 441.5; DB 4; Length 129;

Best Local Similarity 67.4%; Pred. No. 2e-33;

Matches 89; Conservative 10; Mismatches 28; Indels 5; Gaps 2;

QY 1 QVQLDSGGGLVQAGGSLRLSCAASGSIQDIHTWGYRQTPGKQRDVAVTIQSGGSTNYA 60

Db 1 QVQLDSGGGLVQAGGSLRLSCAASGSIQDIHTWGYRQTPGKQKLVATITTYRTNYA 60

QY 61 DSVKGRFTISRDNLTNTLYLQNDLKPEDTGVYNNADVPYRTSRYLEIWGGTTLVT 118

Db 61 DSVKGRFTISRDNLTNTLYLQNDLKPEDTGVYNNADVPYRTSRYLEIWGGTTLVT 117

QY 119 VFLEPKTKPKQP 130

Db 118 VSSEPKTKPKQP 129

Search completed: October 5, 2004, 08:00:14

Job time : 99.9043 secs

This Page Blank (uspc)

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2004, 07:36:16 ; Search time 2.10648 Seconds  
(without alignments)  
593.639 Million cell updates/sec

Title: US-09-805-290A-8

Perfect score: 68

Sequence: 1 DVRPYRTSRYLEV 13

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR\_78:\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	42	61.8	303	2 T47356	hypothetical prote
2	42	61.8	962	1 SNECPI	pitriylisin (EC 3.4
3	42	61.8	962	2 F91088	proteinase III [im
4	42	61.8	962	2 H85933	proteinase III [im
5	42	61.8	962	2 AC0865	protease III precu
6	41	60.3	775	2 E84828	probable WD-40 rep
7	40	58.8	442	2 AD3042	oxidoreductase Atu
8	40	58.8	442	2 H98243	agaE protein [impo
9	40	58.8	594	2 T15202	hypothetical prote
10	39	57.4	258	2 T00447	hypothetical prote
11	39	57.4	526	2 A86274	FA19.15 protein -
12	39	57.4	569	2 B82559	30S ribosomal prot
13	39	57.4	1441	2 T00335	hypothetical prote
14	38	55.9	103	2 T24300	ribosomal protein
15	37	54.4	286	2 T08250	hypothetical prote
16	37	54.4	415	2 G90705	probable transport
17	37	54.4	415	2 B85556	probable transport
18	37	54.4	415	2 G84790	ybdG protein - Esc
19	37	54.4	415	2 A80573	probable membrane
20	37	54.4	437	2 G91067	hypothetical prote
21	37	54.4	441	2 H85911	hypothetical prote
22	37	54.4	599	2 T10798	pherophorin-S - Vo
23	36	52.9	84	2 G86327	protein F18014.23
24	36	52.9	168	2 H72721	hypothetical prote
25	36	52.9	260	2 T00446	hypothetical prote
26	36	52.9	269	2 T16910	hypothetical prote
27	36	52.9	278	2 AG0015	probable membrane
28	36	52.9	310	1 B69475	conserved hypothet
29	36	52.9	376	2 T43846	translation elonga

probable isocitrat  
hypothetical prote  
hypothetical prote  
ribosomal protein  
30S ribosomal prot  
RNA-directed RNA p  
RNA-directed RNA p  
RNA-directed RNA p  
Fas-binding protei  
probable heterochr  
RND multidrug effl  
probable dna-direc  
ataxia telangiecta  
hypothetical prote  
hypothetical prote  
hypothetical prote

30 36 52.9 435 2 A72658  
31 36 52.9 448 2 T48966  
32 36 52.9 536 2 E84665  
33 36 52.9 556 2 B82141  
34 36 52.9 559 2 C83250  
35 36 52.9 725 1 P21VBS  
36 36 52.9 726 1 P21VBC  
37 36 52.9 726 1 P21VBW  
38 36 52.9 740 2 T03847  
39 36 52.9 960 2 T37916  
40 36 52.9 1043 2 H83071  
41 36 52.9 1405 2 T40607  
42 36 52.9 1708 1 A43100  
43 35.5 52.2 787 2 C84898  
44 35 51.5 113 2 C81220  
45 35 51.5 113 2 G81990

#### ALIGNMENTS

##### RESULT 1

T47356  
hypothetical protein F18P9.140 - Arabidopsis thaliana  
C;Species: Arabidopsis thaliana (mouse-ear cress)  
C;Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 20-Apr-2000  
C;Accession: T47356  
R;Yakutara, G.; Partmann, B.; Dauner, D.; Sterr, W.; Holland, R.; Weichselgartner, M.; Mayer, K.F.X.  
Submitted to the Protein Sequence Database, April 2000  
A;Reference number: Z24458  
A;Accession: T47356  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-303 <N>YA>  
A;Cross-references: EMBL:AL138654  
A;Experimental source: cultivar Columbia; BAC clone F18P9  
C;Genetics:  
A;Map position: 3  
A;Introns: 47/3; 103/3; 140/2; 182/1; 210/2; 241/3  
A;Note: F18P9.140

Query Match 61.8%; Score 42; DB 2; Length 303;  
Best Local Similarity 53.8%; Pred. No. 4.9;  
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 DVRPYRTSRYLEV 13

|.|||.|||.:

Db 10 DIRPYKTSWRIQV 22

##### RESULT 2

SNECPI  
pitriylisin (EC 3.4.24.55) precursor [validated] - Escherichia coli (strain K-12)  
N;Alternate names: endopeptidase Pi; proteinase III  
C;Species: Escherichia coli  
C;Date: 31-Mar-1993 #sequence\_revision 31-Oct-1997 #text\_change 27-Oct-2003  
C;Accession: F65064; A29093; A25765; B25532  
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Col  
.A.; Rose, D.J.; Mau, B.; Shao, Y.  
Science 277, 1453-1462, 1997  
A;Title: The complete genome sequence of Escherichia coli K-12.  
A;Reference number: A64720; MUID:97426617; PMID:9278503  
A;Accession: F65064  
A;Status: nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-962 <BLAT>  
A;Cross-references: GB:AE000365; GB:U00096; NID:92367163; PIDN:AACT5860.1; PID:92367164;  
A;Experimental source: Strain K-12, substrain MG1655  
R;Claverie-Martin, F.; Diaz-Torres, M.R.; Kushner, S.R.  
Gene 54, 185-195, 1987  
A;Title: Analysis of the regulatory region of the protease III (ptr) gene of Escherichia  
A;Reference number: A29093; MUID:88005781; PMID:3308636

A/Accession: A29093  
 A/Molecule type: DNA  
 A/Residues: 1-276, 'HVS1R', 283, 'W', 285-296 <CLA>  
 A/Cross-references: GB:M17095; NID:G147390; PIDN:AAA24436.1; PID:G147391  
 A/Experimental source: strain K12  
 A/Note: part of this sequence, including the amino end of the mature protein, was confirmed by R. Finch, P.W.; Wilson, R.E.; Brown, K.; Hickson, I.D.; Emmerson, P.T.  
 A/Title: Complete nucleotide sequence of the *Escherichia coli* ptr gene encoding protease  
 A/Reference number: A25765; MUID:87040734; PMID:3534791  
 A/Accession: A25765  
 A/Molecule type: DNA  
 A/Residues: 1-962 <FIN>  
 A/Cross-references: GB:X06227; NID:G42560; PIDN:CAA95976.1; PID:G42561  
 A/Recker, A.B.; Roth, R.A.  
 A/Title: An unusual active site identified in a family of zinc metalloendopeptidases.  
 A/Reference number: A38854; MUID:92237263; PMID:1570301  
 A/Contents: annotation; active site  
 C/Genetics:  
 A/Gene: ptr  
 A/Map position: 61  
 C/Function:  
 A/Description: endopeptidase degrades small peptides (validated, MUID:92237263)  
 A/Pathway: protein degradation  
 C/Superfamily: insulin-degrading enzyme (IDE)  
 C/Keywords: hydrolase; metalloproteinase; monomer; periplasmic space; protein degradation  
 F/1-23/Domain: signal sequence #status predicted <SIG>  
 F/24-962/Product: pitrilsin #status experimental <KAT>  
 F/88,92/Binding site: zinc (His) #status experimental  
 F/91/Active site: Glu #status experimental

Query Match 61.8%; Score 42; DB 1; Length 962;  
 Best Local Similarity 80.0%; Pred. No. 16;  
 Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 4 PYRTSRYLEV 13  
 ||||: ||||  
 Db 124 PYRTAFYLEV 133

RESULT 3  
 F91088  
 C/Species: *Escherichia coli*  
 C/Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 27-Oct-2003  
 C/Accession: F91088  
 R/Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.; Gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.  
 A/Title: Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genome  
 A/Reference number: A99629; MUID:21156231; PMID:11258796  
 A/Accession: F91088  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-962 <HAY>  
 A/Cross-references: GB:BA000007; PIDN:BA37101.1; PID:G13363150; GSPDB:GN00154  
 A/Experimental source: strain O157:H7, substrain RIMD 0509952  
 C/Genetics:  
 A/Gene: ECs3678  
 C/Superfamily: insulin-degrading enzyme (IDE)

Query Match 61.8%; Score 42; DB 2; Length 962;  
 Best Local Similarity 80.0%; Pred. No. 16;  
 Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 4 PYRTSRYLEV 13  
 ||||: ||||  
 Db 124 PYRTAFYLEV 133

RESULT 4  
 H85933

protease III [imported] - *Escherichia coli* (strain O157:H7, substrain EDL933)  
 C/Species: *Escherichia coli*  
 C/Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 27-Oct-2003  
 C/Accession: H85933  
 R/Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew, M.; Miller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca, Nature 409, 529-533, 2001  
 A/Title: Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7.  
 A/Reference number: A85480; MUID:21074935; PMID:11206551  
 A/Accession: H85933  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-962 <SFO>  
 A/Cross-references: GB:AB005174; NID:G12517303; PIDN:AA57932.1; GSPDB:GN00145; UMGF:Z413  
 A/Experimental source: strain O157:H7, substrain EDL933  
 C/Genetics:  
 A/Gene: ptr  
 C/Superfamily: insulin-degrading enzyme (IDE)

Query Match 61.8%; Score 42; DB 2; Length 962;  
 Best Local Similarity 80.0%; Pred. No. 16;  
 Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 4 PYRTSRYLEV 13  
 ||||: ||||  
 Db 124 PYRTAFYLEV 133

RESULT 5  
 AC0865  
 C/Species: *Salmonella enterica* subsp. *enterica* serovar Typhi  
 A/Note: this species has also been called *Salmonella typhi*  
 C/Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 27-Oct-2003  
 C/Accession: AC0865  
 R/Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moule, S.; O'Gaora, P.  
 A/Title: Complete genome sequence of a multiple drug resistant *Salmonella enterica* serovar Typhi  
 A/Reference number: AB0502; MUID:21534947; PMID:11677608  
 A/Accession: AC0865  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-962 <PAR>  
 A/Cross-references: GB:AL513382; PIDN:CAD02819.1; PID:G16504073; GSPDB:GN00176  
 C/Genetics:  
 A/Gene: STY3133  
 C/Superfamily: insulin-degrading enzyme (IDE)

Query Match 61.8%; Score 42; DB 2; Length 962;  
 Best Local Similarity 80.0%; Pred. No. 16;  
 Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 4 PYRTSRYLEV 13  
 ||||: ||||  
 Db 124 PYRTAFYLEV 133

RESULT 6  
 E84828  
 Probable WD-40 repeat protein [imported] - *Arabidopsis thaliana*  
 C/Species: *Arabidopsis thaliana* (mouse-ear cress)  
 C/Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
 C/Accession: E84828  
 R/Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; M. M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.; euss, D.; Niernan, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J. Nature 402, 761-768, 1999  
 A/Title: Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*.  
 A/Reference number: A84420; MUID:20083487; PMID:10617197  
 A/Accession: E84828

A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-775 <STO>  
A:Cross-references: GB:AE002093; NID:g4586061; PIDN:AAD25679.1; GSPDB:GN00139  
C:Genetics:  
A:Gene: At2g40360  
A:Map position: 2

Query Match 60.3%; Score 41; DB 2; Length 775;  
Best Local Similarity 66.7%; Pred. No. 20;  
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 DVPRYRTSYLYE 12  
| | | | |  
DB 429 DLRYPNRSYLE 440

## RESULT 7

AD3042  
oxidoreductase Atu3952 [imported] - Agrobacterium tumefaciens (strain C58, Dupont)  
C:Species: Agrobacterium tumefaciens  
C:Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 18-Nov-2002  
C:Accession: AD3042  
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.; Karp, P.; Romero, P.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McClellan Science 294, 2317-2323, 2001  
A:Authors: Yoo, H.; Iao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, ster, E.W.  
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.  
A:Reference number: AB2577; MUID:21608550; PMID:11743193  
A:Accession: AD3042  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-442 <KUR>  
A:Cross-references: GB:AE008689; PIDN:AAU44754.1; PID:g17742390; GSPDB:GN00187  
A:Experimental source: strain C58 (Dupont)  
C:Genetics:  
A:Gene: Atu3952  
A:Map position: linear chromosome

Query Match 58.8%; Score 40; DB 2; Length 442;  
Best Local Similarity 58.3%; Pred. No. 17;  
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 DVPRYRTSYLYE 12  
| | | | |  
DB 421 DLRYPNRSRPF 432

## RESULT 8

H98243  
agaE protein [imported] - Agrobacterium tumefaciens (strain C58, Cereon)  
C:Species: Agrobacterium tumefaciens  
C:Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 18-Nov-2002  
C:Accession: H98243  
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman, A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.; Science 294, 2323-2328, 2001  
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tumefaciens  
A:Reference number: A97359; MUID:21608551; PMID:11743194  
A:Accession: H98243  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-442 <KUR>  
A:Cross-references: GB:AE007870; PIDN:AAK89474.1; PID:g15159342; GSPDB:GN00170  
C:Genetics:  
A:Gene: AGR\_L1801  
A:Map position: linear chromosome

Query Match 58.8%; Score 40; DB 2; Length 442;  
Best Local Similarity 58.3%; Pred. No. 17;  
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 DVPRYRTSYLYE 12  
| | | | |  
DB 421 DLRYPNRSRPF 432

## RESULT 9

T15202  
hypothetical protein F12B6.3 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 20-Sep-1999  
C:Accession: T15202  
R:Pauley, A.; Maggi, L.  
submitted to the EMBL Data Library, May 1997  
A:Description: the sequence of C. elegans cosmid F12B6.  
A:Reference number: Z19307  
A:Accession: T15202  
A>Status: preliminary; translated from GB/EMBL/DBDJ  
A:Molecule type: DNA  
A:Residues: 1-594 <PAU>  
A:Cross-references: EMBL:AF003138; NID:g2088708; PID:g2088711; PIDN:AAB54155.1; GSPDB:GN00187  
A:Experimental source: strain Bristol N2; clone F12B6  
C:Genetics:  
A:Gene: CESP:F12B6.3  
A:Map position: 1  
A:Introns: 45/2; 146/3; 211/3; 316/1; 367/3; 407/3; 552/2

Query Match 58.8%; Score 40; DB 2; Length 594;  
Best Local Similarity 61.5%; Pred. No. 23;  
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 DVPRYRTSYLYE 13  
| | | | |  
DB 289 DLSRYRTSHTEV 301

## RESULT 10

T00447  
hypothetical protein T14N5.4 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cross)  
C:Date: 01-Feb-1999 #sequence\_revision 01-Feb-1999 #text\_change 22-Oct-1999  
C:Accession: T00447  
R:Federspiel, N.A.; Palm, C.J.; Conway, A.B.; Kurtz, D.B.; Conway, A.R.; Au, M.; Araujo, V.; Vystotskaia, V.S.; Yu, G.; Ecker, J.; Theologis, A.; Davis, R.W.  
submitted to the EMBL Data Library, September 1998  
A:Reference number: Z14152  
A:Accession: T00447  
A>Status: translated from GB/EMBL/DBDJ  
A:Molecule type: DNA  
A:Residues: 1-258 <PED>  
A:Cross-references: EMBL:AC004260; NID:g3176694; PID:g3540203; GSPDB:GN00059; ATSP:T14N5.  
A:Experimental source: cultivar Columbia  
C:Genetics:  
A:Gene: ATSP:T14N5.4  
A:Map position: 1  
A:Introns: 38/3; 106/1

Query Match 57.4%; Score 39; DB 2; Length 258;  
Best Local Similarity 66.7%; Pred. No. 15;  
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 4 PYRTSYLYE 12  
| | | | |  
DB 199 PWRTSRMYQ 207

## RESULT 11

A86274  
F7A19.15 protein - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cross)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 09-Nov-2001  
C:Accession: A86274  
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,

Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;  
ansen, N.F.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant *Arabidopsis*.  
A:Reference number: A86141; MUID:21016719; PMID:11130712  
A:Accession: A86274  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-526 <STO>  
A:Cross-references: GB:AE005172; NID:G5080782; PIDN:AAD39292.1; GSPDB:GN00141  
C:Genetics:  
A:Map position: 1

Query Match 57.4%; Score 39; DB 2; Length 526;  
Best Local Similarity 87.5%; Pred. No. 31;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 PYRTSYL 11  
|||||  
Db 79 PYRTSYL 86

RESULT 12  
B82559  
30S ribosomal protein S1 XF2438 [imported] - *Xylella fastidiosa* (strain 9a5c)  
C:Species: *Xylella fastidiosa*  
C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Sep-2000  
A:Accession: B82559  
R:Anonymous, The *Xylella fastidiosa* Consortium of the Organization for Nucleotide Sequen  
Nature 406, 151-157, 2000  
A:Title: The genome sequence of the plant pathogen *Xylella fastidiosa*.  
A:Reference number: A82515; MUID:20365717; PMID:10910347  
A:Note: for a complete list of authors see reference number A59328 below  
A:Accession: B82559  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-569 <SIM>  
A:Cross-references: GB:AE004052; GB:AE003849; NID:G9107617; PIDN:AAP95237.1; GSPDB:GN001  
A:Experimental source: strain 9a5c  
R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A  
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H  
as-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.  
submitted to GenBank, June 2000  
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Praga, J.S.; Franca, S.C.; Franco, M.C.; Frohm  
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laig  
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E  
A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;  
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A  
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak  
A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir  
M.; Tsunako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z  
A:Reference number: A59328  
A:Contents: annotation  
C:Genetics:  
A:Gene: XF2438  
C:Superfamily: Escherichia coli ribosomal protein S1

Query Match 57.4%; Score 39; DB 2; Length 569;  
Best Local Similarity 66.7%; Pred. No. 34;  
Matches 8; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 DVPRPRTSYL 12  
|||||  
Db 149 DVPRPRTSYL 160

RESULT 13  
T00335

hypothetical protein KIAA0564 - human (fragment)  
C:Species: Homo sapiens (man)  
C:Date: 01-Feb-1999 #sequence\_revision 01-Feb-1999 #text\_change 21-Jul-2000  
R:Accession: T00335  
R:Nagase, T.; Ishikawa, K.; Miyajima, N.; Tanaka, A.; Kotani, H.; Nomura, N.; Ohara, O.  
DNA Res. 5, 31-39, 1998  
A:Title: Prediction of the coding sequences of unidentified human genes. IX. The complete  
A:Reference number: Z14086; MUID:98290545; PMID:9628581  
A:Accession: T00335  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-1441 <NAG>  
A:Cross-references: EMBL:AB011136; NID:G3043651; PIDN:BAA25490.1; PID:G3043652  
A:Experimental source: brain; clone HH1811  
C:Genetics:  
A:Note: KIAA0564

Query Match 57.4%; Score 39; DB 2; Length 1441;  
Best Local Similarity 61.5%; Pred. No. 87;  
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 DVPRPRTSYL 13  
|||||  
Db 979 DVTPPTSGYIEV 991

RESULT 14  
T24300  
ribosomal protein S14 T01E8.6, mitochondrial [similarity] - *Caenorhabditis elegans*  
C:Species: *Caenorhabditis elegans*  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 03-May-2002  
R:McMurray, A.  
submitted to the EMBL Data Library, March 1995  
A:Reference number: Z19871  
A:Accession: T24300  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-103 <WIL>  
A:Cross-references: EMBL:Z48809; PIDN:CAA88746.1; GSPDB:GN00020; CESP:T01E8.6  
A:Experimental source: clone T01E8  
C:Genetics:  
A:Gene: CESP:T01E8.6  
A:Map position: 2  
A:Introns: 20/3  
C:Superfamily: Escherichia coli ribosomal protein S14

Query Match 55.9%; Score 38; DB 2; Length 103;  
Best Local Similarity 60.0%; Pred. No. 9;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 VPRPRTSYL 11  
:|||||  
Db 76 IKPRLSRHL 85

RESULT 15  
T08250  
hypothetical protein H0411 [imported] - *Halobacterium* sp. (strain NRC-1) plasmid pNRC100  
N:Alternate names: hypothetical protein H1942  
C:Species: *Halobacterium* sp.  
A:Variety: strain NRC-1  
C:Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 03-Nov-2000  
A:Accession: T08250; T08393  
R:Ng, W.V.; Clafo, S.A.; Smith, T.M.; Bungarner, R.E.; Baskin, D.; Faust, J.; Hall, B.; I  
Genome Res. 8, 1131-1141, 1998  
A:Title: Snapshot of a large dynamic replicon in a halophilic Archaeon: megaplasmid or m  
A:Reference number: Z16408; MUID:99063795; PMID:9847077  
A:Accession: T08250  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-286 <NGW>  
A:Cross-references: EMBL:AF016485; NID:G2822278; PID:G2822311; HALOSP:H0411



A:Experimental source: strain NRC-1  
A:Genetics: COP1  
A:Accession: U03333  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-286 <DAS>  
A:Cross-references: EMBL:AF016485; NID:g2822278; PID:g2822454; HALOSP:H1942  
A:Experimental source: strain NRC-1  
A:Genetics: COP2  
C:Genetics: <COP1>  
A:Gene: HALOSP:H0411  
A:Genome: plasmid pNRC100  
C:Genetics: <COP2>  
A:Gene: HALOSP:H1942  
A:Genome: plasmid pNRC100

Query Match 54.4%; Score 37; DB 2; Length 286;  
Best Local Similarity 50.0%; Pred. No. 39;  
Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Qy 1 DVRPYRTSRYLE 12  
|:|:|:|:|  
Db 105 DLQPYRSTRLYE 116

Search completed: October 5, 2004, 08:16:48  
Job time : 7.10648 secs

This Page Blank (uspio)

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2004, 07:29:35 ; Search time 1.16358 Seconds

(without alignments)  
581.749 Million cell updates/sec

Title: US-09-805-290A-8

Perfect score: 58

Sequence: 1 DVPRTSRVLEY 13

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	42	61.8	962	1 PTRA_ECO57	Q8x6m8 escherichia
2	42	61.8	962	1 PTRA_ECOL6	Q8cve2 escherichia
3	42	61.8	962	1 PTRA_ECOLI	P05458 escherichia
4	42	61.8	962	1 PTRA_SALTI	Q82418 salmonella
5	42	61.8	962	1 PTRA_SALTY	Q82mb5 salmonella
6	42	61.8	962	1 PTRA_SHIFL	Q83qc3 shigella fl
7	39	57.4	526	1 FUT7_ARATH	Q9x181 arabidopsis
8	38	55.9	199	1 RT14_CABEL	P49391 caenorhabdi
9	38	55.9	3503	1 DS_DROME	Q24292 drosophila
10	37	54.4	415	1 YBDG_ECOLI	P39455 escherichia
11	37	54.4	620	1 SVR_BIFLO	Q8g4v2 bifidobacte
12	36	52.9	378	1 RRP2_PROSP	P14128 providencia
13	36	52.9	725	1 RRP2_INBSI	P11136 influenza b
14	36	52.9	726	1 RRP2_INBAC	P13873 influenza b
15	36	52.9	726	1 RRP2_INBAD	P13874 influenza b
16	36	52.9	726	1 RRP2_INBP9	Q36432 influenza b
17	36	52.9	731	1 DAXX_RAT	Q8vib2 rattus norv
18	36	52.9	739	1 DAXX_MOUSE	Q35613 mus muscucu
19	36	52.9	740	1 DAXX_HUMAN	Q9uer7 homo sapien
20	36	52.9	960	1 CHPT_SCHPO	Q10103 schizosacch
21	36	52.9	1405	1 RPT1_SCHPO	Q94666 schizosacch
22	36	52.9	3056	1 ATM_HUMAN	P13315 homo sapien
23	36	52.9	3066	1 ATM_MOUSE	Q62388 mus muscucu
24	35.5	52.2	349	1 MOAA_RHIME	Q92pb4 rhizobium m
25	35	51.5	99	1 DEFC_AEDA	P81603 aedes aegypt
26	35	51.5	157	1 LUXS_BORBU	Q50154 borrelia bu
27	35	51.5	251	1 V14F_BPT4	P13321 bacteriopho
28	35	51.5	251	1 V1F_FIVPE	P16089 feline immu
29	35	51.5	251	1 V1F_FIVSD	P19029 feline immu
30	35	51.5	300	1 DHAA_MYCBO	Q9xb14 mycobacteri
31	35	51.5	300	1 DHAA_MYCTU	Q50642 mycobacteri
32	35	51.5	443	1 CBPM_HUMAN	P14384 homo sapien
33	35	51.5	444	1 V111_F0HPV	Q9j5b1 fowlpox vir

34	51.5	487	1 SYE2_THEMA	Q9x2i8 thermotoga
35	51.5	500	1 FUT8_ARATH	Q9x178 arabidopsis
36	51.5	557	1 RSI_ECOLI	P02349 escherichia
37	51.5	566	1 RSI_BUCBP	Q89aj3 buchnera ap
38	51.5	653	1 MALQ_PIRKO	Q32450 pyrococcus
39	51.5	665	1 Z333_HUMAN	Q96j19 homo sapien
40	50.0	132	1 HM01_CABEL	P53547 caenorhabdi
41	50.0	216	1 Y101_ARCFU	O30135 archaeoglob
42	50.0	230	1 FLPA_THEAC	Q9hj18 thermoplasma
43	50.0	282	1 KD12_RHIME	Q92v09 rhizobium m
44	50.0	286	1 DAP2_GLOAB	Q97d80 clostridium
45	50.0	321	1 YSI_CHICK	Q06066 gallus gall

#### ALIGNMENTS

RESULT 1  
PTRA\_SCO57  
ID PTRA\_ECO57 STANDARD; PRT; 962 AA.  
AC Q8X6M8;  
DT 10-OCT-2003 (Rel. 42, Created)  
DT 10-OCT-2003 (Rel. 42, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Protease III precursor (EC 3.4.24.55) (Pitrilysin) (Protease pi).  
GN PTRA OR PTR OR Z4138 OR EC33678.  
OS Escherichia coli O157:H7.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Escherichia.  
OX NCBI\_TaxID=83334;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;  
RX MEDLINE=21074935; PubMed=11206551;  
RA Fern N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,  
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,  
RA Posfai J., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,  
RA Grobeck E.J., Davis N.W., Lim A., Dimantanta E.T., Potamouisis K.,  
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,  
RA Welch R.A., Blattner F.R.;  
RL "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7";  
RT Nature 409:528-533 (2001).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=O157:H7 / RIMD 0509952;  
RX MEDLINE=21156331; PubMed=11258796;  
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,  
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,  
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,  
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;  
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli  
O157:H7 and genomic comparison with a laboratory strain K-12";  
RL DNA Res. 8:11-22 (2001).  
CC -!- FUNCTION: Endopeptidase that degrades small peptides of less than  
CC 7 kDa, such as Glucagon and insulin (By similarity).  
CC -!- CATALYTIC ACTIVITY: Preferential cleavage of 16-Tyr-|-Leu-17 and  
CC 25-Phe-|-Tyr-26 bonds of oxidized insulin B chain. Also acts on  
CC other substrates of Mw less than 7 kDa such as insulin and  
CC glucagon.  
CC -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).  
CC -!- SUBUNIT: Monomer (By similarity).  
CC -!- SUBCELLULAR LOCATION: Periplasmic (By similarity).  
CC -!- SIMILARITY: Belongs to peptidase family M16.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; AB005510; AAC57932.1; --

```
DR EMBL; AP002563; BAB37101.1; -
DR FIR; F91088; F91088. Peptidase M16.
DR InterPro; IPR001431; Peptidase M16.
DR InterPro; IPR007863; Peptidase M16_C.
DR Pfam; PF00675; Peptidase M16; 1.
DR Pfam; PF05193; Peptidase M16 C; 2.
DR PROSITE; PS00143; INSULINASE; 1.
DR Hydrolase; Metalloprotease; Magnesium; Zinc; Signal;
KW Complete proteome.
FT SIGNAL 1 23 BY SIMILARITY.
FT CHAIN 24 962 PROTEASE III.
FT METAL 88 88 ZINC (BY SIMILARITY).
FT ACT_SITE 91 91 ZINC (BY SIMILARITY).
FT METAL 92 92 ZINC (BY SIMILARITY).
FT METAL 169 169 ZINC (BY SIMILARITY).
FT METAL 169 169 ZINC (BY SIMILARITY).
SQ SEQUENCE 962 AA; 107808 MW; 30785644CBF0E452 CRC64;

Query Match 61.8%; Score 42; DB 1; Length 962;
Best Local Similarity 80.0%; Pred. No. 6.2;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 4 PYRTSRYLEV 13
DB 124 PYRTAFYLEV 133

RESULT 2
PTRA_ECOL6
ID PTRA_ECOL6 STANDARD; PRT; 962 AA.
AC Q8CVS2;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Protease III precursor (EC 3.4.24.55) (Pitrilysin) (Protease pi).
GN PTRA OR PTR OR C3415.
OS Escherichia coli O6.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=217992;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=O6:H1 / CFT073 / ATCC 700928;
RX MEDLINE=22388234; PubMed=12471157;
RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
RT "Extensive mosaic structure revealed by the complete genome sequence
of uropathogenic Escherichia coli.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
CC -!- FUNCTION: Endopeptidase that degrades small peptides of less than
CC 7 kDa, such as glucagon and insulin (By similarity).
CC -!- CATALYTIC ACTIVITY: Preferential cleavage of 16-Tyr-[Leu-17 and
CC 25-Phe]-Tyr-26 bonds of oxidized insulin B chain. Also acts on
CC other substrates of Mw less than 7 kDa such as insulin and
CC glucagon.
CC -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC -!- SUBUNIT: Monomer (By similarity).
CC -!- SUBCELLULAR LOCATION: Periplasmic (By similarity).
CC -!- SIMILARITY: Belongs to peptidase family M16.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB016765; AAN81860.1; -
DR InterPro; IPR001431; Peptidase M16.
DR InterPro; IPR007863; Peptidase M16_C.
DR Pfam; PF00675; Peptidase M16; 1.

DR Pfam; PF05193; Peptidase M16 C; 2.
DR PROSITE; PS00143; INSULINASE; 1.
DR Hydrolase; Metalloprotease; Magnesium; Zinc; Signal;
KW Complete proteome.
FT SIGNAL 1 23 BY SIMILARITY.
FT CHAIN 24 962 PROTEASE III.
FT METAL 88 88 ZINC (BY SIMILARITY).
FT ACT_SITE 91 91 ZINC (BY SIMILARITY).
FT METAL 92 92 ZINC (BY SIMILARITY).
FT METAL 169 169 ZINC (BY SIMILARITY).
FT METAL 169 169 ZINC (BY SIMILARITY).
SQ SEQUENCE 962 AA; 107808 MW; 30785644CBF0E452 CRC64;

Query Match 61.8%; Score 42; DB 1; Length 962;
Best Local Similarity 80.0%; Pred. No. 6.2;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 4 PYRTSRYLEV 13
DB 124 PYRTAFYLEV 133

RESULT 3
PTRA_ECOLI
ID PTRA_ECOLI STANDARD; PRT; 962 AA.
AC P05458; P78106;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-NOV-1988 (Rel. 09, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Protease III precursor (EC 3.4.24.55) (Pitrilysin) (Protease pi).
GN PTRA OR PTR OR B2821.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
SEQUENCE FROM N.A.
RC MEDLINE=87040734; PubMed=534791;
RX Finch P.W., Wilson R.B., Brown K., Hickson I.D., Emerson P.T.;
RA "Complete nucleotide sequence of the Escherichia coli ptr gene
encoding protease III.";
RL Nucleic Acids Res. 14:7695-7703(1986).
RN [2]
SEQUENCE FROM N.A.
RC STRAIN=K12 / MGL655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [3]
SEQUENCE OF 853-962 FROM N.A.
RX MEDLINE=87066729; PubMed=3537960;
RA Finch P.W., Storey A., Chapman K.E., Brown K., Hickson I.D.,
RA Emerson P.T.;
RT "Complete nucleotide sequence of the Escherichia coli recB gene.";
RL Nucleic Acids Res. 14:8573-8582(1986).
RN [4]
SEQUENCE OF 1-296 FROM N.A.
RC STRAIN=K12;
RX MEDLINE=88005781; PubMed=3308636;
RA Claverie-Martin F., Diaz-Torres M.R., Kushner S.R.;
RT "Analysis of the regulatory region of the protease III (ptr) gene of
Escherichia coli K-12.";
RL Gene 54:185-195(1987).
RN [5]
MUTAGENESIS, AND ACTIVE SITE.
RX MEDLINE=92237263; PubMed=1570301;
RA Becker A.B., Roth R.A.;
RT "An unusual active site identified in a family of zinc
metalloendopeptidases.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:3835-3839(1992).
```

[6]  
RN MUTAGENESIS, AND ACTIVE SITE.  
RX MEDLINE=93277449; PubMed=8092978;  
RA Becker A.B., Roth R.A.;  
RT Identification of glutamate-169 as the third zinc-binding residue in  
RT proteinase III, a member of the family of insulin-degrading  
RT enzymes.";  
RL Biochem. J. 292:137-142(1993).  
CC -!- FUNCTION: Endopeptidase that degrades small peptides of less than  
CC 7 kDa, such as glucagon and insulin.  
CC -!- CATALYTIC ACTIVITY: Preferential cleavage of 16-Tyr-|-Leu-17 and  
CC 25-Phe-|-Tyr-26 bonds of oxidized insulin B chain. Also acts on  
CC other substrates of Mw less than 7 kDa such as insulin and  
CC glucagon.  
CC -!- COFACTOR: Binds 1 zinc ion per subunit.  
CC -!- SUBUNIT: Monomer.  
CC -!- SUBCELLULAR LOCATION: Periplasmic.  
CC -!- SIMILARITY: Belongs to peptidase family M16.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; X04581; CAA28249.1; -;  
DR EMBL; U29581; AAB40468.1; -;  
DR EMBL; AE000365; AAC75860.1; -;  
DR EMBL; X06227; CAA29576.1; -;  
DR EMBL; M17095; AAA24436.1; -;  
DR PIR; P65064; SNECFI.  
DR MEROPS; M16.001; -;  
DR EC2DBASE; G095.0; 6TH EDITION.  
DR EcoGene; EGI0786; ptrA.  
DR InterPro; IPR001431; Peptidase\_M16.  
DR Pfam; PF00675; Peptidase\_M16\_C.  
DR Pfam; PF05193; Peptidase\_M16\_I\_2.  
DR PROSITE; PS00143; INSULINASE; 1.  
KW Hydrolase; Metalloprotease; Magnesium; Periplasmic; Zinc; Signal;  
Complete proteome.  
FT SIGNAL 1 23  
FT CHAIN 24 962 PROTEASE III.  
FT METAL 88 88 ZINC.  
FT ACT\_SITE 91 91  
FT METAL 92 92 ZINC.  
FT METAL 169 169 ZINC.  
FT MUTAGEN 88 88 H-R: LOSS OF ACTIVITY AND OF ZN-BINDING.  
FT MUTAGEN 91 91 E-S: LOSS OF ACTIVITY.  
FT MUTAGEN 92 92 H-R: LOSS OF ACTIVITY AND OF ZN-BINDING.  
FT MUTAGEN 162 162 E-S: 20% LOSS OF ACTIVITY.  
FT MUTAGEN 169 169 E-S: LOSS OF ACTIVITY AND OF ZN-BINDING.  
FT MUTAGEN 204 204 E-S: NO LOSS OF ACTIVITY.  
FT CONFLICT 277 284 IHHVFA -> HYHSLRPW (IN REF. 4).  
SQ SEQUENCE 962 AA; 107708 MW; 0558C68C2F1A0540 CRC64;  
  
Query Match 61.8%; Score 42; DB 1; Length 962;  
Best Local Similarity 80.0%; Pred. No. 6.2;  
Matches 8; Conservative 1; Mismatches 0; Gaps 0;  
  
QY 4 PYRTSRYLEV 13  
Db 124 PYRTAFYLEV 133  
  
RESULT 4  
PTRA SALT  
ID PTRA SALT STANDARD; PRT; 962 AA.  
AC Q82418;  
DT 10-OCT-2003 (Rel. 42, Created)  
DT 10-OCT-2003 (Rel. 42, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Protease III precursor (BC 3.4.24.55) (Pitrilysin) (Protease pi).  
GN PTRA OR PTR OR STV3133 OR T2903.  
OS Salmonella typhi.  
CC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
CC Enterobacteriaceae; Salmonella.  
OX NCBI\_TaxID=601;  
RN [1]  
RN SEQUENCE FROM N.A.  
RC STRAIN=CT18;  
RC MEDLINE=21534947; PubMed=11677608;  
RX Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,  
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebaihia M.,  
RA Baker S., Basham D., Brooks K., Chillingworth T., Conerton P.,  
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,  
RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagals K.,  
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,  
RA Quail M.A., Rutherford K., Simmonds M., Skelton J., Stevens K.,  
RA Whitehead S., Barrrell B.G.;  
RT "Complete genome sequence of a multiple drug resistant Salmonella  
RT enterica serovar Typhi CT18.";  
RL Nature 413:848-852(2001).  
[2]  
RN SEQUENCE FROM N.A.  
RC STRAIN=Ty2 / ATCC 700931;  
RC MEDLINE=22531367; PubMed=12644504;  
RX Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,  
RA Burland V., Kodoyianni V., Schwartz D.C., Blattner F.R.;  
RT "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2  
RT and CT18.";  
RL J. Bacteriol. 185:2330-2337(2003).  
CC -!- FUNCTION: Endopeptidase that degrades small peptides of less than  
CC 7 kDa, such as glucagon and insulin (By similarity).  
CC -!- CATALYTIC ACTIVITY: Preferential cleavage of 16-Tyr-|-Leu-17 and  
CC 25-Phe-|-Tyr-26 bonds of oxidized insulin B chain. Also acts on  
CC other substrates of Mw less than 7 kDa such as insulin and  
CC glucagon.  
CC -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).  
CC -!- SUBUNIT: Monomer (By similarity).  
CC -!- SUBCELLULAR LOCATION: Periplasmic (By similarity).  
CC -!- SIMILARITY: Belongs to peptidase family M16.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; AL627277; CAD02819.1; -;  
DR EMBL; AE016843; AAO70457.1; -;  
DR MEROPS; M16.001; -;  
DR InterPro; IPR001431; Peptidase\_M16.  
DR InterPro; IPR007863; Peptidase\_M16\_C.  
DR Pfam; PF00675; Peptidase\_M16\_I\_2.  
DR Pfam; PF05193; Peptidase\_M16\_C\_2.  
DR PROSITE; PS00143; INSULINASE; FALSE NEG  
DR Hydrolase; Metalloprotease; Magnesium; Periplasmic; Zinc; Signal;  
KW Complete proteome.  
FT SIGNAL 1 23 BY SIMILARITY.  
FT CHAIN 24 962 PROTEASE III.  
FT METAL 88 88 ZINC (BY SIMILARITY).  
FT ACT\_SITE 91 91 BY SIMILARITY.  
FT METAL 92 92 ZINC (BY SIMILARITY).  
FT METAL 169 169 ZINC (BY SIMILARITY).  
SQ SEQUENCE 962 AA; 107524 MW; 5745C2C77F7F5832 CRC64;  
  
Query Match 61.8%; Score 42; DB 1; Length 962;  
Best Local Similarity 80.0%; Pred. No. 6.2;  
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
  
QY 4 PYRTSRYLEV 13

```

Db      124 PYRTAFYLEV 133
||||: ||||
RESULT 5
PTRA SALTY
ID PTRA SALTY STANDARD; PRT; 962 AA.
AC Q8ZME5,
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DE Protease III precursor (EC 3.4.24.55) (Pitrilysin) (Protease pi).
GN PTRA OR PTR OR STM2955.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E.; Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du P., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2";
RL Nature 413:852-856(2001).
CC -!- FUNCTION: Endopeptidase that degrades small peptides of less than
CC 7 kDa, such as glucagon and insulin (By similarity).
CC -!- CATALYTIC ACTIVITY: Preferential cleavage of 16-Tyr-|-Leu-17 and
CC 25-Phe-|-Tyr-26 bonds of oxidized insulin B chain. Also acts on
CC other substrates of Mw less than 7 kDa such as insulin and
CC glucagon.
CC -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC -!- SUBUNIT: Monomer (By similarity).
CC -!- SUBCELLULAR LOCATION: Periplasmic (By similarity).
CC -!- SIMILARITY: Belongs to peptidase family M16.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE008837; AAL21871.1; -.
CC StyGene; SG27777; PTRa.
CC InterPro; IPR001431; Peptidase_M16.
CC InterPro; IPR007863; Peptidase_M16_C.
CC Pfam; PF00675; Peptidase_M16; 1.
CC Pfam; PF05193; Peptidase_M16_C; 2.
CC PROSITE; PS00143; INSULINASE; 1.
CC Hydrolase; Metalloprotease; Magnesium; Periplasmic; Zinc; Signal;
KW Complete proteome.
FT SIGNAL 1 23 BY SIMILARITY.
FT CHAIN 24 962 PROTEASE III.
FT METAL 88 88 ZINC (BY SIMILARITY).
FT ACT_SITE 91 91 ZINC (BY SIMILARITY).
FT METAL 92 92 ZINC (BY SIMILARITY).
FT METAL 169 169 ZINC (BY SIMILARITY).
FT SEQUENCE 962 AA; 107486 MW; 322AD6E87B873952 CRC64;
Query Match 61.8%; Score 42; DB 1; Length 962;
Best Local Similarity 80.0%; Pred. No. 6.2;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 4 PYRTSRYLEV 13
||||: ||||
Db      124 PYRTAFYLEV 133
RESULT 6
PTRA SHIFL
ID PTRA SHIFL STANDARD; PRT; 962 AA.
AC Q83QC3;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DE Protease III precursor (EC 3.4.24.55) (Pitrilysin) (Protease pi).
GN PTRA OR PTR OR SF2832 OR S3029.
OS Shigella flexneri.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Shigella.
OX NCBI_TaxID=623;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=301 / Serotype 2a;
RX MEDLINE=22272406; PubMed=12384590;
RA Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,
RA Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,
RA Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,
RA Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,
RA Yu J.;
RT "Genome sequence of Shigella flexneri 2a: insights into pathogenicity
RT through comparison with genomes of Escherichia coli K12 and O157";
RL Nucleic Acids Res. 30:4432-4441(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=2457T / ATCC 700930 / Serotype 2a;
RX MEDLINE=22590274; PubMed=12704152;
RA Wei J., Goldberg M.B., Burland V., Venkatesan M.M., Deng W.,
RA Fournier G., Mayhew G.F., Plunkett G. III, Rose D.J., Darling A.,
RA Mau B., Perna N.T., Payne S.M., Runyen-Janecky L.J., Zhou S.,
RA Schwartz D.C., Blattner F.R.;
RT "Complete genome sequence and comparative genomics of Shigella
RT flexneri serotype 2a strain 2457T";
RL Infect. Immun. 71:2775-2786(2003).
CC -!- FUNCTION: Endopeptidase that degrades small peptides of less than
CC 7 kDa, such as glucagon and insulin (By similarity).
CC -!- CATALYTIC ACTIVITY: Preferential cleavage of 16-Tyr-|-Leu-17 and
CC 25-Phe-|-Tyr-26 bonds of oxidized insulin B chain. Also acts on
CC other substrates of Mw less than 7 kDa such as insulin and
CC glucagon.
CC -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC -!- SUBUNIT: Monomer (By similarity).
CC -!- SUBCELLULAR LOCATION: Periplasmic (By similarity).
CC -!- SIMILARITY: Belongs to peptidase family M16.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE015297; AAN44319.1; -.
CC EMBL; AE016987; AAP18144.1; -.
CC InterPro; IPR001431; Peptidase_M16.
CC InterPro; IPR007863; Peptidase_M16_C.
CC Pfam; PF00675; Peptidase_M16; 1.
CC Pfam; PF05193; Peptidase_M16_C; 2.
CC PROSITE; PS00143; INSULINASE; 1.
CC Hydrolase; Metalloprotease; Magnesium; Periplasmic; Zinc; Signal;
KW Complete proteome.
FT SIGNAL 1 23 BY SIMILARITY.
FT CHAIN 24 962 PROTEASE III.
FT METAL 88 88 ZINC (BY SIMILARITY).
FT ACT_SITE 91 91 ZINC (BY SIMILARITY).
FT METAL 92 92 ZINC (BY SIMILARITY).
FT METAL 169 169 ZINC (BY SIMILARITY).
FT SEQUENCE 962 AA; 107770 MW; 662041A5DCA0F254 CRC64;

```

```

DR EMBL; AC007576; AAD39292.1; -.
DR EMBL; AC068197; AAF79408.1; ALT_SEQ.
DR PIR; A86274; A86274.
DR InterPro; IPRO04938; XG F1ase.
DR Pfam; PF03254; XG F1ase; 1.
DR Transfase; Glycosyltransferase; Transmembrane; Glycoprotein;
DR Signal-anchor; Golgi stack; Cell wall.
DR DOMAIN 1 4 CYTOPLASMIC (POTENTIAL).
DR TRANSMEM 5 25 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
(POTENTIAL).
DR DOMAIN 26 526 LUMENAL, CATALYTIC (POTENTIAL).
DR CARBOHYD 211 211 N-LINKED (GLCNAC. .) (POTENTIAL).
DR CARBOHYD 215 215 N-LINKED (GLCNAC. .) (POTENTIAL).
DR CARBOHYD 363 363 N-LINKED (GLCNAC. .) (POTENTIAL).
DR SEQUENCE 526 AA; 60537 MW; 8F3B447551025FCD CRC64;

Query Match 57.4%; Score 39; DB 1; Length 526;
Best Local Similarity 87.5%; Pred. No. 12;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 PYRTSYL 11
DB 79 PYRTSEYL 86
|||||

RESULT 8
RT14_CAEEL
IID RT14 CAEEL STANDARD; PRT; 199 AA.
AC P49391;
AC AC
DT 01-FEB-1996 (Rel. 33, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Putative mitochondrial 40S ribosomal protein S14.
DE T0188.6
OS Caenorhabditis elegans.
OS Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
NCBI_TaxID=6239;
[1]_
OX OX
RN SEQUENCE FROM N.A.
RC STRAIN-Bristol N2;
RA McMurray A.A.;
RL Submitted (MAR-1995) to the EMBL/GenBank/DDBJ databases.
RN [2]
RN REVISIONS.
RA Durbin R.;
RL Submitted (DEC-2002) to the EMBL/GenBank/DDBJ databases.
CC -!- SUBCELLULAR LOCATION: Mitochondrial (By similarity).
CC -!- SIMILARITY: Belongs to the S14p family of ribosomal proteins.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z48809; CA88746.2; -.
DR PIR; T24300; T24300.
DR WormPep; T01E8.6; CE32923.
DR InterPro; IPRO01209; Ribosomal_S14.
DR Pfam; PF0253; Ribosomal_S14; 1.
DR PROSITE; PS00527; RIBOSOMAL_S14; FALSE_NEG.
DR Ribosomal protein; Mitochondrion.
DR SEQUENCE 199 AA; 23414 MW; 2CC0021A618BCFF0 CRC64;

Query Match 55.9%; Score 38; DB 1; Length 199;
Best Local Similarity 60.0%; Pred. No. 6.4;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 VRPYRTSYL 11
|||||

```

DB 172 IKPYRLSRHL 181

RESULT 9  
DS\_DROME STANDARD; PRT; 3503 AA.

AC Q24292; Q9VPS4;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Dacheous protein precursor (Adherin).  
GN DS CR C017941.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]\_TaxID=7227;  
RP SEQUENCE FROM N.A., FUNCTION, TISSUE SPECIFICITY, AND DEVELOPMENTAL STAGE.  
RP TISSUE=Embryo;  
RC MEDLINE=95324813; PubMed=7601355;  
RA Clark H.F., Brentrup D., Schnetz K., Bieber A., Goodman C., Noll M.;  
RT "Dachsous encodes a member of the cadherin superfamily that controls imaginal disc morphogenesis in Drosophila.";  
RL Genes Dev. 9:1530-1542(1995).  
[2]  
RP REVISIONS.  
RP Noll M.;  
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.  
[3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Berkley;  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galie R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
RA Brill J.F., Agbayani A., An H.-J., Andrews-Frankkoch C., Baldwin D.,  
RA Ballaw R.M., Basu A., Bakendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlike C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Flooker C., Gabrielian A.B., Garg N.S., Gelbart W.M., Glaeser K.,  
RA Gladek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimbel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Laske P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.R., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Sider-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wassman D.A., Weinstock G.M., Weissbach J.,  
RA Williams S.M., Woodage T., Worley J.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-P., Zaveri J.S., Zhan M., Zhang G., Zhao X., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of Drosophila melanogaster."  
RN Science 287:2185-2195(2000).  
[4]

RP REVISIONS.  
RP MEDLINE=22426069; PubMed=12537572;  
RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,  
RA Hradscky P., Huang Y., Kaminker J.S., Millburn G.H., Prochuk S.E.,  
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,  
RA Bettencourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,  
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,  
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,  
RA Lewis S.E.;  
RT "Annotation of the Drosophila melanogaster euchromatic genome: a systematic review";  
RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).  
CC -1- FUNCTION: Involved in morphogenesis. May also be involved in cell adhesion.  
CC -1- SUBCELLULAR LOCATION: Type I membrane protein (Potential).  
CC -1- TISSUE SPECIFICITY: Expressed in embryonic ectoderm in larvae, expression is restricted to imaginal disks and brain.  
CC -1- DEVELOPMENTAL STAGE: Expressed throughout embryogenesis where it is first detected during gastrulation. Also expressed in larvae and adults.  
CC -1- SIMILARITY: Contains 27 cadherin domains.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
EMBL; L08811; AAA79329.2; ;  
EMBL; AE003588; AAF51468.3; ALT\_INIT.  
HSP; P15116; INCU  
FlyBase; FBgn000497; ds.  
GO; GO:0005887; C:integral to plasma membrane; ISS.  
GO; GO:0008014; F:calcium-dependent cell adhesion molecule ac.; ISS.  
GO; GO:0016339; F:calcium-dependent cell-cell adhesion; ISS.  
GO; GO:0008283; P:cell proliferation; IMP.  
GO; GO:0009094; P:cellular morphogenesis during differentiation; IMP.  
GO; GO:0045317; P:equator specification; IMP.  
GO; GO:0045398; P:establishment of epithelial cell polarity; IMP.  
GO; GO:0018149; P:protein-protein cross-linking; IPI.  
InterPro; IPR002126; Cadherin.  
InterPro; IPR002233; Cadherin\_C\_term.  
Pfam; PF01049; Cadherin; 26.  
Pfam; PF01049; Cadherin\_C term; 1.  
PRINTS; PR00205; CADHERIN.  
SMART; SMO0112; CA; 25.  
PROSITE; PS00232; CADHERIN\_1; 20.  
PROSITE; PS0268; CADHERIN\_2; 27.  
Cell adhesion; Glycoprotein; Transmembrane; Calcium; Calcium-binding; Repeat; Signal; Developmental protein.  
SIGNAL 1 20  
POTENTIAL.  
FT CHAIN 21 3503 DACHSUS PROTEIN  
FT DOMAIN 21 3045 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 3046 3066 POTENTIAL.  
FT DOMAIN 3067 3503 CYTOPLASMIC (POTENTIAL).  
FT DOMAIN 22 121 CADHERIN 1.  
FT DOMAIN 122 233 CADHERIN 2.  
FT DOMAIN 234 340 CADHERIN 3.  
FT DOMAIN 345 451 CADHERIN 4.  
FT DOMAIN 452 558 CADHERIN 5.  
FT DOMAIN 559 662 CADHERIN 6.  
FT DOMAIN 663 774 CADHERIN 7.  
FT DOMAIN 775 878 CADHERIN 8.  
FT DOMAIN 879 983 CADHERIN 9.  
FT DOMAIN 984 1100 CADHERIN 10.  
FT DOMAIN 1101 1203 CADHERIN 11.  
FT DOMAIN 1205 1312 CADHERIN 12.  
FT DOMAIN 1313 1432 CADHERIN 13.  
FT DOMAIN 1433 1549 CADHERIN 14.  
FT DOMAIN 1556 1666 CADHERIN 15.  
FT DOMAIN 1667 1794 CADHERIN 16.



```

FT DOMAIN 1796 1899 CADHERIN 17.
FT DOMAIN 1900 2004 CADHERIN 18.
FT DOMAIN 2005 2111 CADHERIN 19.
FT DOMAIN 2114 2269 CADHERIN 20.
FT DOMAIN 2270 2375 CADHERIN 21.
FT DOMAIN 2375 2479 CADHERIN 22.
FT DOMAIN 2489 2595 CADHERIN 23.
FT DOMAIN 2596 2699 CADHERIN 24.
FT DOMAIN 2701 2809 CADHERIN 25.
FT DOMAIN 2810 2916 CADHERIN 26.
FT DOMAIN 2919 3028 CADHERIN 27.
FT CONFLICT 1070 1070 V -> I (IN REF. 1).
FT CONFLICT 1490 1490 R -> S (IN REF. 1).
FT CONFLICT 1636 1636 G -> S (IN REF. 1).
FT CONFLICT 1692 1692 S -> P (IN REF. 1).
FT CONFLICT 1804 1804 V -> I (IN REF. 1).
FT CONFLICT 2029 2029 L -> I (IN REF. 1).
FT CONFLICT 2210 2210 P -> A (IN REF. 1).
FT CONFLICT 2289 2289 A -> S (IN REF. 1).
FT CONFLICT 2536 2536 S -> T (IN REF. 1).
FT CONFLICT 2862 2862 R -> Q (IN REF. 1).
FT CONFLICT 3038 3038 S -> G (IN REF. 1).
SQ SEQUENCE 3503 AA; 379774 MW; 975B09F059F7EEF5 CRC64;

Query Match 55.9%; Score 38; DB 1; Length 3503;
Best Local Similarity 60.0%; Pred. No. 1.4e+02;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 DVPYRTSRY 10
Db 149 DLEPYNTORY 158

RESULT 10
ID YBDG ECOLI STANDARD; PRT; 415 AA.
AC P39455; P77602;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical protein ybdg.
GN YBDG OR B0577 OR SF0483 OR S0492.
OS Escherichia coli, and
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562, 623;
[1]
RN SEQUENCE FROM N.A.
RC SPECIES=E.coli; STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden W.A., Rose D.J.,
RA Mau B., Shao Y.;
RA "The complete genome sequence of Escherichia coli K-12."
RT Science 277:1453-1474 (1997).
[2]
RN SEQUENCE FROM N.A.
RC SPECIES=E.coli; STRAIN=K12 / MG1655;
RA Chung E., Allen E., Araujo R., Aparicio A., Davis K., Duncan M.,
RA Federapfel N., Hyman R., Kallman S., Komp C., Kuri O., Lew H., Lin D.,
RA Namath A., Osifner P., Roberts D., Schramm S., Davis R.W.;
RA Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
[3]
RN SEQUENCE FROM N.A.
RC SPECIES=E.coli; STRAIN=K12;
RX MEDLINE=97061202; PubMed=8905232;
RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA Mori H., Motomura K., Nakamura Y., Nashimoto H., Nishio Y., Saito N.,
RA Sempel G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,

```

```

RA Yano M., Horiuchi T.;
RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 12.7-28.0 min region on the linkage map.";
RL DNA Res. 3:1137-1155 (1996).
[4]
RN SEQUENCE OF 1-77 FROM N.A.
RC SPECIES=E.coli; STRAIN=B;
RX MEDLINE=95113294; PubMed=7813889;
RA Michael N.P., Brehm J.K., Anlezark G.M., Minton N.P.;
RT "Physical characterisation of the Escherichia coli B gene encoding
RT nitroreductase and its over-expression in Escherichia coli K12.";
RL FEMS Microbiol. Lett. 124:195-202 (1994).
[5]
RN SEQUENCE OF 1-12 FROM N.A.
RC SPECIES=E.coli; STRAIN=K12 / C600;
RA Zenno S., Koike H., Tanokura M., Saigo K.;
RL Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.
[6]
RN SEQUENCE FROM N.A.
RC SPECIES=S.flexneri; STRAIN=301 / Serotype 2a;
RX MEDLINE=22272406; PubMed=12384590;
RA Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,
RA Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,
RA Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,
RA Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,
RA Yu J.;
RT "Genome sequence of Shigella flexneri 2a: insights into pathogenicity
RT through comparison with genomes of Escherichia coli K12 and O157.";
RL Nucleic Acids Res. 30:4432-4441 (2002).
[7]
RN SEQUENCE FROM N.A.
RC SPECIES=S.flexneri; STRAIN=2457T / ATCC 700930 / Serotype 2a;
RX MEDLINE=22590274; PubMed=12704152;
RA Wei J., Goldberg M.B., Burland V., Venkatesan M.M., Deng W.,
RA Fournier G., Mayhew G.F., Plunkett G. III, Rose D.J., Darling A.,
RA Mau B., Perna N.T., Payne S.M., Runyen-Janecky L.J., Zhou S.,
RA Schwartz D.C., Blattner F.R.;
RT "Complete genome sequence and comparative genomics of Shigella
RT flexneri serotype 2a strain 2457T.";
RL Infect. Immun. 71:2775-2786 (2003).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AB001631; AAC73678.1; -
CC EMBL; U82598; AAB40775.1; -
CC EMBL; D90700; BAA35217.1; -
CC EMBL; U07860; AAC43265.1; -
CC EMBL; D25414; -; NOT ANNOTATED CDS.
CC EMBL; AB015080; AAN42134.1; ALT_INIT.
CC EMBL; AE016979; AAP16008.1; -
CC EMBL; G64790; G64790.
CC EcoGene; EGI2620; ybdg.
CC InterPro; IPR006685; MSion channel.
CC Pfam; PF00924; MS channel; 1.
KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 25 45 POTENTIAL.
FT TRANSMEM 68 88 POTENTIAL.
FT TRANSMEM 105 125 POTENTIAL.
FT TRANSMEM 149 169 POTENTIAL.
FT TRANSMEM 174 194 POTENTIAL.
FT TRANSMEM 370 390 POTENTIAL.
FT TRANSMEM 415 AA; 46602 MW; C9A8A8A41CDC8F28 CRC64;
SQ SEQUENCE 415 AA; 46602 MW; C9A8A8A41CDC8F28 CRC64;

Query Match 54.4%; Score 37; DB 1; Length 415;
Best Local Similarity 50.0%; Pred. No. 22;
Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

```

```

QY      2 VRPRTSRVLE 13
DB      291 LRPLYTSRQEI 302

RESULT 11
SYR_BIFLO
ID SYR_BIFLO STANDARD; PRT; 620 AA.
AC QGAV2;
DT 15-MAR-2004 (Rel. 43, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Arginyl-tRNA synthetase (EC 6.1.1.19) (Arginine--tRNA ligase) (ArgRS).
GN ARG5 OR BL1272.
OS Bifidobacterium longum.
OC Bacteria; Actinobacteria; Actinobacteridae; Bifidobacteriales;
OC Bifidobacteriaceae; Bifidobacterium.
OX NCBI_TaxID=216816;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCC 2705;
RX MEDLINE=2294977; PubMed=12381787;
RA Scheil M.A., Karmirantzou M., Snel B., Vilanova D., Berger B.,
RA Pessi G., Zwahlen M.-C., Desiere F., Bork P., Delley M.,
RA Priddy R.D., Arigoni F.;
RT "The genome sequence of Bifidobacterium longum reflects its adaptation
RT to the human gastrointestinal tract.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:14422-14427(2002)..
CC -!- CATALYTIC ACTIVITY: ATP + L-arginine + tRNA(Arg) = AMP +
CC diphosphate + L-arginyl-tRNA(Arg).
CC -!- SUBUNIT: Monomer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AB014754; AAN25073.1; -.
CC HAMAP; MF_00123; -.
CC InterPro; IPR001278; Arg_tRNA-synt_1c.
CC InterPro; IPR005148; N.
CC InterPro; IPR008909; tRNA-synt_id_c.
CC InterPro; IPR001412; tRNA-synt_id_c.
CC InterPro; IPR009080; tRNA-synt_id_bind.
CC Pfam; PF03485; N-Arg; 1.
CC Pfam; PF00750; tRNA-synt_id; 1.
CC Pfam; PF05746; tRNA-synt_id_c; 1.
CC PRINTS; PR01038; TRNASYNTHARG.
CC TIGRFAMs; TIGR00456; argS; 1.
CC PROSITE; PS00178; AA tRNA LIGASE I; 1.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW Complete proteome.
FT SITE 147 157 "HIGH" REGION.
FT SEQUENCE 620 AA; 67206 MW; F51C3C0C8CD9AB06 CRC64;

Query Match 54.4%; Score 37; DB 1; Length 620;
Best Local Similarity 50.0%; Pred. No. 34;
Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY      1 DVRPRTSRVLE 12
DB      534 DRQPHKVARYLE 545

RESULT 12
RSP_PROSP
ID RSP_PROSP STANDARD; PRT; 378 AA.
AC QGAV2;
DT 15-MAR-2004 (Rel. 43, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Arginyl-tRNA synthetase (EC 6.1.1.19) (Arginine--tRNA ligase) (ArgRS).
GN ARG5 OR BL1272.
OS Bifidobacterium longum.
OC Bacteria; Actinobacteria; Actinobacteridae; Bifidobacteriales;
OC Bifidobacteriaceae; Bifidobacterium.
OX NCBI_TaxID=216816;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCC 2705;
RX MEDLINE=2294977; PubMed=12381787;
RA Scheil M.A., Karmirantzou M., Snel B., Vilanova D., Berger B.,
RA Pessi G., Zwahlen M.-C., Desiere F., Bork P., Delley M.,
RA Priddy R.D., Arigoni F.;
RT "The genome sequence of Bifidobacterium longum reflects its adaptation
RT to the human gastrointestinal tract.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:14422-14427(2002)..
CC -!- CATALYTIC ACTIVITY: ATP + L-arginine + tRNA(Arg) = AMP +
CC diphosphate + L-arginyl-tRNA(Arg).
CC -!- SUBUNIT: Monomer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AB014754; AAN25073.1; -.
CC HAMAP; MF_00123; -.
CC InterPro; IPR001278; Arg_tRNA-synt_1c.
CC InterPro; IPR005148; N.
CC InterPro; IPR008909; tRNA-synt_id_c.
CC InterPro; IPR001412; tRNA-synt_id_c.
CC InterPro; IPR009080; tRNA-synt_id_bind.
CC Pfam; PF03485; N-Arg; 1.
CC Pfam; PF00750; tRNA-synt_id; 1.
CC Pfam; PF05746; tRNA-synt_id_c; 1.
CC PRINTS; PR01038; TRNASYNTHARG.
CC TIGRFAMs; TIGR00456; argS; 1.
CC PROSITE; PS00178; AA tRNA LIGASE I; 1.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW Complete proteome.
FT SITE 147 157 "HIGH" REGION.
FT SEQUENCE 620 AA; 67206 MW; F51C3C0C8CD9AB06 CRC64;

Query Match 54.4%; Score 37; DB 1; Length 620;
Best Local Similarity 50.0%; Pred. No. 34;
Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY      1 DVRPRTSRVLE 12
DB      534 DRQPHKVARYLE 545

RESULT 13
RRP2_INBSI
ID RRP2_INBSI STANDARD; PRT; 725 AA.
AC P11136;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE RNA-directed RNA polymerase subunit P2 (EC 2.7.7.48) (Polymerase
DE acidic protein) (PA).
OS Influenza B virus (strain B/Singapore/222/79).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenzavirus B.
OX NCBI_TaxID=11544;
RN [1]
RP SEQUENCE FROM N.A.

```

```

AC P14128;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 30S ribosomal protein S1 (Fragment).
GN RPSA.
OS Providencia sp.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Providencia.
OX NCBI_TaxID=589;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86013618; PubMed=3862933;
RA Schnier J., Faist G.;
RT "Comparative studies on the structural gene for the ribosomal protein
RT S1 in ten bacterial species.";
RL Mol. Gen. Genet. 200:476-481(1995).
CC -!- FUNCTION: BINDS MRNA, THUS FACILITATING RECOGNITION OF THE
CC INITIATION POINT. IT IS NEEDED TO TRANSLATE MRNA WITH A SHORT
CC SHINE-DALGARNO (SD) PURINE-RICH SEQUENCE (BY SIMILARITY).
CC -!- SIMILARITY: Belongs to the S1P family of ribosomal proteins.
CC -!- SIMILARITY: Contains 5 S1 motif domains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X02828; CAA36596.1; -.
CC HSPP; P05055; LSRO.
CC InterPro; IPR008994; Nucleic acid OB.
CC InterPro; IPR000110; Ribosomal_S1.
CC InterPro; IPR003029; S1.
CC Pfam; PF00575; S1; 5.
CC PRINTS; PR00681; RIBOSOMALS1.
CC SMART; SM00316; S1; 4.
CC PROSITE; PS0126; S1; 5.
KW Ribosomal protein; Repeat; RNA-binding.
FT NON_TER 1 1
FT DOMAIN <1 66 S1 MOTIF 1.
FT DOMAIN 87 155 S1 MOTIF 2.
FT DOMAIN 172 242 S1 MOTIF 3.
FT DOMAIN 259 329 S1 MOTIF 4.
FT DOMAIN 346 378 S1 MOTIF 5.
FT SEQUENCE 378 AA; 41613 MW; 9052424346D2E904 CRC64;

Query Match 52.9%; Score 36; DB 1; Length 378;
Best Local Similarity 58.3%; Pred. No. 31;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY      1 DVRPRTSRVLE 12
DB      32 DVRPVRTTHLE 43

RESULT 13
RRP2_INBSI
ID RRP2_INBSI STANDARD; PRT; 725 AA.
AC P11136;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE RNA-directed RNA polymerase subunit P2 (EC 2.7.7.48) (Polymerase
DE acidic protein) (PA).
OS Influenza B virus (strain B/Singapore/222/79).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenzavirus B.
OX NCBI_TaxID=11544;
RN [1]
RP SEQUENCE FROM N.A.

```

```

DR PFam: PF00603; Flu_PA; 1.
DR Transfaser: RNA-directed RNA polymerase.
DR SEQ SEQUENCE 726 AA; 33181 MW; CAFD1A4388034F26 CRC64;

Query Match 52.9%; Score 36; DB 1; Length 726;
Best Local Similarity 46.2%; Pred. No. 62;
Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 DVPYRTSYRLEV 13
   | : | : | : | : |
DB 109 DLFYKTKRFIEV 121

RESULT 15
RRP2_INBP2
ID_RRP2_INBP2 STANDARD; PRT; 726 AA.
AC P13874;
01-JAN-1990 (Rel. 13, Created)
01-JAN-1990 (Rel. 13, Last sequence update)
28-FEB-2003 (Rel. 41, Last annotation update)
RNA-directed RNA polymerase subunit P2 (EC 2.7.7.48) (Polymerase
acidic protein) (PA).
DE Influenza B virus (strain B/Ann Arbor/1/66 [wild-type]).
OS Influenza B virus (strain B/Ann Arbor/1/66 [wild-type]).
OS Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenzavirus B.
OC NCBI_TaxID=11523;
[1]
RRN SEQUENCE FROM N.A.
RRP MEDLINE=88179548; PubMed=3354202;
RX Deborde D.C.; Donabedian A.M.; Herlocher M.L.; Naeve C.W.;
RA Maesbø H.F.;
RA "Sequence comparison of wild-type and cold-adapted B/Ann Arbor/1/66
RT Influenza virus genes.";
RT Virology 163:429-443(1988).
RRL -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC {RNA}(N).
CC -!- SUBUNIT: Influenza RNA polymerase is composed of three subunits:
CC P1 (or PB1), P2 (or PA), and P3 (or PB2).
CC -!- SIMILARITY: Belongs to the influenza viruses polymerase PA family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M20172; AAA43766.1; -.
CC FIR; F28604; P2IVBM
CC InterPro; IPR001009; RNA_pol_P2.
CC PFam; PF00603; Flu_PA; 1.
CC Transfaser: RNA-directed RNA polymerase.
CC SEQ SEQUENCE 726 AA; 83166 MW; C9D85DA44844019E CRC64;

Query Match 52.9%; Score 36; DB 1; Length 726;
Best Local Similarity 46.2%; Pred. No. 62;
Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 DVPYRTSYRLEV 13
   | : | : | : | : |
DB 109 DLFYKTKRFIEV 121

Search completed: October 5, 2004, 08:01:37
Job time : 4.16358 secs

```



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model  
Run on: October 5, 2004, 07:34:11 ; Search time 6.88117 Seconds  
(without alignments)  
596.081 Million cell updates/sec

Title: US-09-805-290A-8  
Perfect score: 68  
Sequence: 1 DVAPYRTSRYLEV 13

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues  
Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

- Database : SPTREMBL 25:\*
- 1: sp\_archaea:\*
  - 2: sp\_bacteria:\*
  - 3: sp\_fungi:\*
  - 4: sp\_human:\*
  - 5: sp\_invertebrate:\*
  - 6: sp\_mammal:\*
  - 7: sp\_mhc:\*
  - 8: sp\_organelle:\*
  - 9: sp\_phase:\*
  - 10: sp\_plant:\*
  - 11: sp\_rodent:\*
  - 12: sp\_virus:\*
  - 13: sp\_vertebrate:\*
  - 14: sp\_unclassified:\*
  - 15: sp\_rvirus:\*
  - 16: sp\_bacteriap:\*
  - 17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	42	61.8	303	10 Q9MIK7	Q9MIK7 arabidopsis
2	42	61.8	581	10 Q9LTU1	Q9LTU1 arabidopsis
3	41	60.3	451	11 Q35390	Q35390 rattus norv
4	41	60.3	775	10 Q9SIY9	Q9SIY9 arabidopsis
5	40	58.8	272	10 Q8LAQ8	Q8LAQ8 arabidopsis
6	40	58.8	442	16 Q8U8Y6	Q8U8Y6 agrobacteri
7	40	58.8	509	5 Q01792	Q01792 caenorhabdi
8	40	58.8	654	2 Q52001	Q52001 streptococc
9	40	58.8	766	5 P91590	P91590 ciona intes
10	39	57.4	196	2 Q7WZ84	Q7WZ84 nonomurae
11	39	57.4	217	5 Q25368	Q25368 loligo opal
12	39	57.4	258	10 Q80652	Q80652 arabidopsis
13	39	57.4	276	16 Q884I1	Q884I1 pseudomonas
14	39	57.4	323	17 Q8TGZ6	Q8TGZ6 methanopyru
15	39	57.4	462	11 Q8CON9	Q8CON9 mus musculu
16	39	57.4	558	16 Q87BJ7	Q87BJ7 xylella fas

17	39	57.4	559	16 Q8P8P5	Q8P8P5 xanthomonas
18	39	57.4	561	16 Q8PK77	Q8PK77 xanthomonas
19	39	57.4	569	16 Q8PAQ7	Q8PAQ7 xylella fas
20	39	57.4	726	12 Q3QLJ3	Q3QLJ3 influenza b
21	39	57.4	980	10 Q84SF5	Q84SF5 oryza sativ
22	39	57.4	1441	4 Q60310	Q60310 homo sapien
23	38	55.9	51	16 Q7UW68	Q7UW68 rhodospirell
24	38	55.9	197	2 Q84AM7	Q84AM7 streptomyce
25	38	55.9	250	10 Q8LR50	Q8LR50 oryza sativ
26	38	55.9	254	5 Q9VJY1	Q9VJY1 drosophila
27	38	55.9	254	5 Q8MZ93	Q8MZ93 drosophila
28	38	55.9	270	16 Q81VM1	Q81VM1 bacillus an
29	38	55.9	270	16 Q811Z5	Q811Z5 bacillus ce
30	38	55.9	322	16 Q7WGG8	Q7WGG8 bordetella
31	38	55.9	322	16 Q7W4Z0	Q7W4Z0 bordetella
32	38	55.9	417	16 Q92RU9	Q92RU9 rhizobium m
33	38	55.9	422	17 Q974N2	Q974N2 sulfolobus
34	38	55.9	463	16 Q7VYQ2	Q7VYQ2 bordetella
35	38	55.9	465	16 Q7W995	Q7W995 bordetella
36	38	55.9	2185	9 Q8W6J4	Q8W6J4 sinorhizobi
37	37	54.4	136	16 Q8PF48	Q8PF48 xanthomonas
38	37	54.4	286	17 Q54583	Q54583 halobacteri
39	37	54.4	358	12 Q8B5N1	Q8B5N1 wisteria ve
40	37	54.4	371	16 Q83FN0	Q83FN0 tropheryma
41	37	54.4	415	16 Q8ZRS2	Q8ZRS2 salmonella
42	37	54.4	415	16 Q8XBK4	Q8XBK4 escherichia
43	37	54.4	415	16 Q8FK34	Q8FK34 escherichia
44	37	54.4	415	16 Q8Z8M7	Q8Z8M7 salmonella
45	37	54.4	441	16 Q8X972	Q8X972 escherichia

ALIGNMENTS

RESULT 1  
Q9MIK7 PRELIMINARY; PRT; 303 AA.  
AC Q9MIK7; (TREMELrel. 15, Created)  
DT 01-OCT-2000 (TREMELrel. 15, Last sequence update)  
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)  
DE Hypothetical protein.  
GN F18P9.140.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.  
CX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Nyakatura G., Partmann B., Dauner D., Sterr W., Holland R.,  
RA Weichselgartner M., Mewes H.W., Rudd S., Lemcke K., Mayer K.F.X.,  
RA Quettier F., Salanoubat M.;  
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA EU Arabidopsis sequencing project;  
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.  
DR ENBL; ALI38654; CAB86685.1; -.  
DR PIR; T47356; T47356.  
DR InterPro; IPR003871; DUF223.  
DR InterPro; IPR008994; Nucleic\_acid\_OB.  
DR Pfam; PF02721; DUF223; 1.  
KW Hypothetical protein.  
SQ SEQUENCE 303 AA; 34913 MW; 5FPE5C14628BA2E7 CRC64;

Query Match 61.8%; Score 42; DB 10; Length 303;  
Best Local Similarity 53.8%; Pred. No. 14;  
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;  
QY 1 DVAPYRTSRYLEV 13  
Db 10 DIRPYKTSWRIOQ 22

```

RESULT 2
Q9LTU1 ID Q9LTU1 PRELIMINARY; PRT; 581 AA.
AC Q9LTU1; 2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 25, Last annotation update)
DE Replication protein A1-like.
OS Arabidopsis thaliana (Mouse-ear cross).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC euroids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Columbia;
RA Sato S., Nakamura Y., Kaneko T., Kato T., Asamizu E., Tabata S.;
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Columbia;
RA Nakamura Y.;
RX MEDLINE=20277480; PubMed=10819329;
RA "Structural analysis of Arabidopsis thaliana chromosome 3. I. Sequence
RT features of the regions of 4,504,864 bp covered by sixty P1 and TAC
RT clones.";
RL DNA Res. 7:131-135(2000).
DR EMBL; AB024034; BAB02796.1; -.
DR InterPro; IPR003871; DUF223.
DR InterPro; IPR008994; Nucleic_acid_OB.
DR Pfam; PF02721; DUF223.1.
DR PFam; PF02721; DUF223.1.
SQ SEQUENCE 581 AA; 64867 MW; 5BCB4DB5B0E006DC CRC64;

Query Match 61.8%; Score 42; DB 10; Length 581;
Best Local Similarity 61.5%; Pred. No. 28;
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 DVEPYRTSYLEV 13
DB 10 DVEPYRTSYLEV 22
|||||:|:|:|

RESULT 3
Q35390 ID Q35390 PRELIMINARY; PRT; 451 AA.
AC Q35390; 1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE Endo-alpha-D-mannosidase.
GN ENMAN
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley;
RX MEDLINE=98030822; PubMed=9361017;
RA Spiro M.J., Bhoyroo V.D., Spiro R.G.;
RT "Molecular cloning and expression of rat liver endo-alpha-mannosidase,
RT an N-linked oligosaccharide processing enzyme.";
RL J. Biol. Chem. 272:29356-29363(1997).
DR EMBL; AF023657; AAB86925.1; -.
DR EMBL; AF023657; AAB86925.1; -.
SQ SEQUENCE 451 AA; 51671 MW; 5A987BFD07585066 CRC64;

Query Match 60.3%; Score 41; DB 11; Length 451;
Best Local Similarity 53.8%; Pred. No. 33;
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 DVEPYRTSYLEV 13

RESULT 4
Q9SIY9 ID Q9SIY9 PRELIMINARY; PRT; 775 AA.
AC Q9SIY9; 2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Putative WD-40 repeat protein.
GN AT2G40360.
OS Arabidopsis thaliana (Mouse-ear cross).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC euroids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=20083487; PubMed=10617197;
RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,
RA Buell C.R., Ketchum K.A., Lee J.J., Renning C.M., Koo H., Moffat K.S.,
RA Cronin L.A., Shen M., VanAken S.E., Umayam L., Tallon L.J., Gill J.E.,
RA Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
RA Copenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,
RA Salzberg S.L., Fraser C.M., Venter J.C.;
RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis
RT thaliana.";
RL Nature 402:761-768(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Lin X.;
RX Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
EMBL; AC007020; AAD25679.1; -.
DR F1R; E84828; E84828.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 4.
DR SMART; SM00320; WD40; 4.
DR PROSITE; PS00678; WD_REPEATS_1; 1.
DR PROSITE; PS00082; WD_REPEATS_2; 1.
DR PROSITE; PS0294; WD_REPEATS_REGION; 2.
KW Repeat; WD repeat.
SQ SEQUENCE 775 AA; 88051 MW; CBF6CF54B528F4EA CRC64;

Query Match 60.3%; Score 41; DB 10; Length 775;
Best Local Similarity 66.7%; Pred. No. 58;
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 DVEPYRTSYLEV 12
DB 429 DVEPYRTSYLEV 440
|||||:|:|:|

RESULT 5
Q8LAQ8 ID Q8LAQ8 PRELIMINARY; PRT; 272 AA.
AC Q8LAQ8; 2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DE Hypothetical protein.
OS Arabidopsis thaliana (Mouse-ear cross).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC euroids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Haas B.J., Volfovsky N., Town C.D., Troukhan M., Alexandrov N.,

```

RA Feldmann K.A., Flavell R.B., White O., Salzberg S.L.;  
RT "Full-length messenger RNA sequences greatly improve genome  
RL annotation."; Genome Biol. 0:0-0(2002).  
RN [2]

RP SEQUENCE FROM N.A.  
RA Brover V., Troukhan M., Alexandrov N., Lu Y.-P., Flavell R.,  
RA Feldmann K., Flavell R.B., White O., Salzberg S.L.;  
RT "Full-length cDNA from Arabidopsis thaliana";  
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AY087671; AAM65208.1; -  
DR InterPro; IPR005162; Retrotrans\_gag.  
DR Pfam; PF03732; Retrotrans\_gag; 1.  
KW Hypothetical protein.  
SQ SEQUENCE 272 AA; 30811 MW; 9C889BD760029D09 CRC64;

Query Match 58.8%; Score 40; DB 10; Length 272;  
Best Local Similarity 46.2%; Pred. No. 30;  
Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 DVRPYRTSRYLEV 13  
DB 97 DTHPRMSRYIQL 109

RESULT 6  
Q8USY6 PRELIMINARY; PRT; 442 AA.  
ID Q8USY6  
AC Q8USY6  
DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Oxidoreductase.  
GN ATU3952 OR AGR L 1801.  
OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
OC Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.  
OX NCBI\_TaxID=176299;  
RN [1]

RP SEQUENCE FROM N.A.  
RX MEDLINE=21608550; PubMed=11743193;  
RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,  
RA Okura Y.K., Zhou Y., Chen L.J., Wood G.E., Almeida N.F. Jr., Woo L.,  
RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Bovee D. Sr.,  
RA Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,  
RA Kutayav T., Levy R., Li M.-J., McClelland E., Palmieri A.,  
RA Raymond C., Rouse G., Saenphammachak C., Wu Z., Romero P., Gordon D.,  
RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan M., Perry M.,  
RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,  
RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,  
RA Nester E.W.;  
RT "The genome of the natural genetic engineer Agrobacterium tumefaciens  
C58.";  
RL Science 294:2317-2323(2001).  
RN [2]

RP SEQUENCE FROM N.A.  
RX MEDLINE=21608551; PubMed=11743194;  
RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,  
RA Qurollo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,  
RA Houmlel K., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu F.,  
RA Wollam C., Allinger M., Doughty D., Scott C., Lappas C., Markelz B.,  
RA Flanagan C., Crowell C., Gursun J., Lomo C., Sear C., Strub G.,  
RA Cielo C., Slater S.;  
RT "Genome sequence of the plant pathogen and biotechnology agent  
RT Agrobacterium tumefaciens C58.";  
RL Science 294:2323-2328(2001).  
DR EMBL; AE009326; AAL44754.1; -  
DR EMBL; AE008288; AAK89474.1; -  
DR PIR; AD3042; AD3042.  
DR PIR; H98243; H98243.  
DR GO; GO:0016491; F:oxidoreductase activity; IEA.  
DR GO; GO:0006118; P:electron transport; IEA.  
DR InterPro; IPR006076; Fad\_oxred.

DR InterPro; IPR00205; NAD\_BS.  
DR Pfam; PF01266; DAO; 1.  
KW Complete proteome.  
SQ SEQUENCE 442 AA; 48309 MW; B074DB4BCFFB62D CRC64;

Query Match 58.8%; Score 40; DB 16; Length 442;  
Best Local Similarity 58.3%; Pred. No. 49;  
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 DVRPYRTSRYLEV 12  
DB 421 DPRPYRSRFFD 432

RESULT 7  
O01792 PRELIMINARY; PRT; 509 AA.  
ID O01792  
AC O01792  
DT 01-JUL-1997 (TrEMBLrel. 04, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Hypothetical protein.  
GN F12B6.3.

OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RX MEDLINE=99069613; PubMed=9851916;  
RA None;  
RT "Genome sequence of the nematode C. elegans: a platform for  
RL Science 282:2012-2018(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RA Pauley A., Maggi L.;  
RT "The sequence of C. elegans cosmid F12B6.";  
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RA Waterston R.;  
RT "Direct Submission.";  
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.  
CC 1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).  
CC 1- SIMILARITY: BELONGS TO THE LIGAND-GATED IONIC CHANNEL FAMILY.  
DR EMBL; AF003138; AAK21371.2; -

DR WormPep; F12B6.3; CE29757.  
DR GO; GO:0016021; C:integral to membrane; IEA.  
DR GO; GO:0005230; F:extracellular ligand-gated ion channel acti. .; IEA.  
DR GO; GO:0005216; F:ion channel activity; IEA.  
DR GO; GO:0030594; P:neurotransmitter receptor activity; IEA.  
DR GO; GO:0005811; P:ion transport; IEA.  
DR GO; GO:0007268; P:synaptic transmission; IEA.  
DR InterPro; IPR006201; Neur\_chan.  
DR InterPro; IPR006202; Neur\_chan\_LBD.  
DR InterPro; IPR006029; Neur\_chan\_memb.  
DR Pfam; PF02931; Neur\_chan\_LBD; 1.  
DR Pfam; PF02932; Neur\_chan\_memb; 1.  
DR PRINTS; PR00252; NRIONCHANNEL.  
DR PROSITE; PS00336; NEUROTR\_ICN\_CHANNEL; 1.  
KW Hypothetical protein; Glycoprotein; Ionic channel;  
KW Postsynaptic membrane; Transmembrane.  
SQ SEQUENCE 509 AA; 58820 MW; 5B86CCA26E90A263 CRC64;

Query Match 58.8%; Score 40; DB 5; Length 509;  
Best Local Similarity 61.5%; Pred. No. 57;  
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 DVRPYRTSRYLEV 13

```

Db      204 DLSRYTSRHEV 216
      |: |||||: ||
      |: |||||: ||

RESULT 8
Q52001 PRELIMINARY; PRT; 654 AA.
AC Q52001;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE NICKASE.
OS Streptococcus sp., and
OS Enterococcus faecalis (Streptococcus faecalis).
OG Plasmid pIP501, and plasmid pRE25.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1306, 1351;
RN [1]
RN SEQUENCE FROM N.A.
RC SPECIES=Streptococcus sp.; STRAIN=FVAL1702; PLASMID=pIP501;
RX MEDLINE=95362640; PubMed=7635806;
RA Wang A., Macrina F.L.;
RT "Streptococcal plasmid pIP501 has a functional orit site.";
RL J. Bacteriol. 177:4199-4206(1995).
RN [2]
RN SEQUENCE FROM N.A.
RC SPECIES=E.faecalis; STRAIN=RE25; PLASMID=pRE25;
RX PubMed=11735367;
RA Schwarz F.V., Perreten V., Teuber M.;
RT "Sequence of the 50-kb conjugative multiresistance plasmid pRE25 from
RT Enterococcus faecalis RE25.";
RL Plasmid 46:170-187(2001).
DR EMBL; L39769; AAA99466.1; -
DR EMBL; X92945; CAC29179.1; -
DR GO; GO:0046821; C:extrachromosomal DNA; IEA.
DR GO; GO:0009291; P:unidirectional conjugation; IEA.
DR InterPro; IPR005053; MobA_MobL.
DR Pfam; PF03389; MobA_MobL; 1.
KW Plasmid.
SQ SEQUENCE 654 AA; 76493 MW; C3EE5B8BB4468600 CRC64;

Query Match 58.8%; Score 40; DB 2; Length 654;
Best Local Similarity 60.0%; Pred. No. 74;
Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 3 RPYRTSRYLE 12
      |: |||||:
      |: |||||:
Db      418 KPFTSRYLE 427

RESULT 9
P91590 PRELIMINARY; PRT; 766 AA.
AC P91590;
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Transglutininase.
OS Ciona intestinalis.
OC Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
OC Phlebobranchia; Clonidae; Ciona.
OX NCBI_TaxID=7719;
RN [1]
RN SEQUENCE FROM N.A.
RX MEDLINE=97330926; PubMed=9187361;
RA Carriello L., Ristoratore P., Zanetti L.;
RT "A new transglutaminase-like from the ascidian Ciona intestinalis.";
RL FEBS Lett. 408:1171-1176(1997).
RN [2]
RN SEQUENCE FROM N.A.
RA Carriello L.;
RT Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.

```

```

DR EMBL; Y10212; CAA71263.1; -
DR HSP; P00488; ICGU.
DR InterPro; IPR001102; Glutransfg.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR008958; Transglut_C.
DR InterPro; IPR002931; Trnsglutase_1_like.
DR Pfam; PF00927; Transglutamin_C; 1.
DR Pfam; PF00868; Transglutamin_N; 1.
DR Pfam; PF01841; Transglut_core; 1.
DR SMART; SMO0460; TGC; 1.
SQ SEQUENCE 766 AA; 87194 MW; 7F90D921D87FDD1D CRC64;

Query Match 58.8%; Score 40; DB 5; Length 766;
Best Local Similarity 77.8%; Pred. No. 88;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DVRPYRTSR 9
      |: |||||: ||
      |: |||||: ||
Db      732 EVRPYRSSR 740

RESULT 10
Q7WZ84 PRELIMINARY; PRT; 196 AA.
AC Q7WZ84;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Putative Vany-type carboxypeptidase.
GN DRV7.
OS Nonomuraea sp. ATCC 39727.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptosporangineae; Streptosporangiaceae; Nonomuraea.
OX NCBI_TaxID=93944;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=ATCC 39727;
RA Sosio M., Stinchi S., Beltrametti F., Lazzarini A., Donadio S.;
RT "The gene cluster for the biosynthesis of the glycopeptide antibiotic
RT A40926 by Nonomuraea sp.";
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ561198; CAD91202.1; -
DR EMBL; AJ561198; CAD91202.1; -
KW Acyltransferase; Carboxypeptidase; Glycosyltransferase; Monoxygenase.
SQ SEQUENCE 196 AA; 22157 MW; 003AA44D19127982 CRC64;

Query Match 57.4%; Score 39; DB 2; Length 196;
Best Local Similarity 58.3%; Pred. No. 32;
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 1 DVRPYRTSRYLE 12
      |: |||||: ||
      |: |||||: ||
Db      137 DVRFEGARWLE 148

RESULT 11
Q25368 PRELIMINARY; PRT; 217 AA.
AC Q25368;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE S-crystallin.
OS Loligo opalescens (California market squid).
OC Eukaryota; Metazoa; Mollusca; Cephalopoda; Coleoidea; Neocoleoidea;
OC Decapodiformes; Loliginidae; Loligo.
OX NCBI_TaxID=31211;
RN [1]
RN SEQUENCE FROM N.A.
RP TISSUE=Lens;
RX MEDLINE=96139039; PubMed=8587103;
RA Tomarev S.I., Chung S., Platiogorsky J.;
RT "Glutathione S-transferase and S-crystallins of cephalopods: evolution
RT from active enzyme to lens-refractive proteins.";

```



RC STRAIN=DC3000;  
RA Buell R., Joardar V., Khouri H., Fedorova N., Tran B., Russell D.,  
RA Berry K., Utterback T., Van Aken S., Feldblyum T., Gwinn M.,  
RA Dodson R., DeBoy R., Durkin A., Kolonay J., Madupu R., Daugherty S.,  
RA Brinkman L., Beanan M., Haft D., Selengut J., Nelson W., Davidsen T.,  
RA White O., Fraser C., Collmer A.,  
RT "Complete sequence of *Pseudomonas syringae*."  
RT Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AE016863; AAC55629.1; -  
DR TIGR; PSPT02112; -  
DR GO; GO:0003934; P:GTP cyclohydrolase I activity; IEA.  
DR GO; GO:0016787; P:hydrolase activity; IEA.  
DR GO; GO:0003058; P:biosynthesis; IEA.  
DR InterPro; IPR001474; GTP\_cyclohydrol.  
DR Pfam; PF01227; GTP\_cyclohydrol; 1.  
KW Hydrolase; Complete proteome.  
SQ SEQUENCE 276 AA; 30751 MW; 693A2CCACBE7D9D2 CRC64;

Query Match 57.4%; Score 39; DB 16; Length 276;  
Best Local Similarity 46.2%; Pred. No. 46;  
Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 DVRYPTSRYLEV 13  
|:|:|:|:|:  
Db 256 DNPYRSTVLDV 268

RESULT 14  
Q8TQZ6 PRELIMINARY; PRT; 323 AA.  
ID Q8TQZ6 AC Q8TQZ6  
DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE tRNA/rRNA cytosine-C5-methylase.  
GN MK0370.  
OS Methanopyrus kandleri.  
OC Archaea; Euryarchaeota; Methanopyri; Methanopyrales; Methanopyraceae;  
OC Methanopyrus.  
OX NCBI\_TaxID=2320;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=AV19 / DSM 6324 / JCM 9639;  
RX MEDLINE=21927647; PubMed=11930014;  
RA Slesarev A.I., Mezhevaya K.V., Makarova K.S., Polushin N.N.,  
RA Shcherbinina O.V., Shakhova V.V., Belova G.I., Aravind L.,  
RA Natale D.A., Rogozin I.B., Tatusov R.L., Wolf Y.I., Stetter K.O.,  
RA Malykh A.G., Koonin E.V., Kozlovskiy S.A.,  
RT "The complete genome of hyperthermophile Methanopyrus kandleri AV19  
and the phylogeny of archaeal methanogens."  
RL Proc. Natl. Acad. Sci. U.S.A. 99:4644-4649 (2002).  
DR EMBL; AE010334; AA001585.1; -  
DR GO; GO:0008757; P:S-adenosylmethionine-dependent methyltransf.; IEA.  
DR InterPro; IPR000051; SAM\_Bind.  
DR InterPro; IPR001678; Sun\_NopI/Nop2.  
DR Pfam; PF01189; Noll\_Nop2\_Sun; 1.  
DR TIGRfams; TIGR00446; nop2p; 1.  
DR PROSITE; PS01153; NOLL\_NOP2\_SUN; 1.  
KW Methyltransferase; Complete proteome.  
SQ SEQUENCE 323 AA; 36336 MW; 32AC13F1B5646A98 CRC64;

Query Match 57.4%; Score 39; DB 17; Length 323;  
Best Local Similarity 70.0%; Pred. No. 55;  
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 VRPYRTSRYL 11  
|:|:|:|:|:  
Db 19 VRPYWARYL 28

RESULT 15  
Q8QCN9 PRELIMINARY; PRT; 462 AA.  
ID Q8QCN9

J. Mol. Evol. 41:1048-1056(1995).  
EMBL; U19300; AAA97551.1; -  
DR HSSP; P46088; 2GSO.  
DR InterPro; IPR004046; GST\_Cterm.  
DR InterPro; IPR004045; GST\_Nterm.  
DR InterPro; IPR003083; S\_crystallin.  
DR Pfam; PF00043; GST\_C; 1.  
DR Pfam; PF02798; GST\_N; 1.  
DR PRINTS; PR01269; SCRSTALLIN.  
SQ SEQUENCE 217 AA; 26113 MW; 66D49D79D9B36F8A CRC64;

Query Match 57.4%; Score 39; DB 5; Length 217;  
Best Local Similarity 50.0%; Pred. No. 36;  
Matches 5; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 DVRYPTSRYLE 12  
|:|:|:|:|:  
Db 125 DLPYMQTRYME 136

RESULT 12  
Q8Q652 PRELIMINARY; PRT; 258 AA.  
ID Q8Q652 AC Q8Q652  
DT 01-NOV-1998 (TrEMBLrel. 08, Created)  
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE T14N5.4 protein.  
GN T14N5.4  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Columbia;  
RA Federspiel N.A., Palm C.J., Conway A.B., Kurtz D.B., Conway A.R.,  
RA Au M., Araujo J., Buehler E., Dewar K., Feng J., Kim C., Li Y.,  
RA Oji O., Osborne B.I., Shinn P., Sun H., Toriumi M., Vysotskaia V.S.,  
RA Yu G., Ecker J., Theologis A., Davis R.W.;  
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AC004260; AAC34353.1; -  
DR F1R; T00447; T00447.  
DR InterPro; IPR006502; DUF\_plant\_1615.  
DR Pfam; PF04720; DUF506; 1.  
DR TIGRfams; TIGR01615; A.thal.3542; 1.  
SQ SEQUENCE 258 AA; 29578 MW; FBC341434DFA6380 CRC64;

Query Match 57.4%; Score 39; DB 10; Length 258;  
Best Local Similarity 66.7%; Pred. No. 43;  
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 4 PYRTSRYLE 12  
|:|:|:|:|:  
Db 199 PWRTSRMQ 207

RESULT 13  
Q88411 PRELIMINARY; PRT; 276 AA.  
ID Q88411 AC Q88411  
DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE GTP cyclohydrolase I, putative.  
GN PSPT02112.  
OS Pseudomonas syringae (pv. tomato).  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
OC Pseudomonadaceae; Pseudomonas.  
OX NCBI\_TaxID=323;  
RN [1]  
RP SEQUENCE FROM N.A.

RC STRAIN=DC3000;  
RA Buell R., Joardar V., Khouri H., Fedorova N., Tran B., Russell D.,  
RA Berry K., Utterback T., Van Aken S., Feldblyum T., Gwinn M.,  
RA Dodson R., DeBoy R., Durkin A., Kolonay J., Madupu R., Daugherty S.,  
RA Brinkman L., Beanan M., Haft D., Selengut J., Nelson W., Davidson T.,  
RA White O., Fraser C., Collmer A.,  
RT "Complete sequence of *Pseudomonas syringae*,"  
RT Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AE016863; AAC55629.1; -  
DR TIGR; PSPT02112; -  
DR GO; GO:0003934; P:GTP cyclohydrolase I activity; IEA.  
DR GO; GO:0016787; P:hydrolase activity; IEA.  
DR GO; GO:0003058; P:biosynthesis; IEA.  
DR InterPro; IPR001474; GTP\_cyclohydrol.  
DR Pfam; PF01227; GTP\_cyclohydrol; 1.  
KW Hydrolase; Complete proteome.  
SQ SEQUENCE 276 AA; 30751 MW; 693A2CCACBE7D9D2 CRC64;

Query Match 57.4%; Score 39; DB 16; Length 276;  
Best Local Similarity 46.2%; Pred. No. 46;  
Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 DVRYPTSRYLEV 13  
|:|:|:|:|:  
Db 256 DNPYRSTVLDV 268

RESULT 14  
Q8TQZ6 PRELIMINARY; PRT; 323 AA.  
ID Q8TQZ6 AC Q8TQZ6; 21, Created)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE tRNA/rRNA cytosine-C5-methylase.  
GN MK0370.  
OS Methanopyrus kandleri.  
OC Archaea; Euryarchaeota; Methanopyri; Methanopyrales; Methanopyraceae;  
OC Methanopyrus.  
OX NCBI\_TaxID=2320;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=AV19 / DSM 6324 / JCM 9639;  
RX MEDLINE=21927647; PubMed=11930014;  
RA Slesarev A.I., Mezhevaya K.V., Makarova K.S., Polushin N.N.,  
RA Shcherbinina O.V., Shakhova V.V., Belova G.I., Aravind L.,  
RA Natale D.A., Rogozin I.B., Tatusov R.L., Wolf Y.I., Stetter K.O.,  
RA Malykh A.G., Koonin E.V., Kozlovskiy S.A.,  
RT "The complete genome of hyperthermophile Methanopyrus kandleri AV19  
and the phylogeny of archaeal methanogens,"  
RL Proc. Natl. Acad. Sci. U.S.A. 99:4644-4649 (2002).  
DR EMBL; AE010334; AA001585.1; -  
DR GO; GO:0008757; P:S-adenosylmethionine-dependent methyltransf.; IEA.  
DR InterPro; IPR000051; SAM\_Bind.  
DR InterPro; IPR001678; Sun\_NopI/Nop2.  
DR Pfam; PF01189; Noll\_Nop2\_Sun; 1.  
DR TIGRfams; TIGR00446; nop2p; 1.  
DR PROSITE; PS01153; NOLL\_NOP2\_SUN; 1.  
KW Methyltransferase; Complete proteome.  
SQ SEQUENCE 323 AA; 36336 MW; 32AC13F1B5646A98 CRC64;

Query Match 57.4%; Score 39; DB 17; Length 323;  
Best Local Similarity 70.0%; Pred. No. 55;  
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 VRPYRTSRYL 11  
|:|:|:|:|:  
Db 19 VRPYWARYL 28

RESULT 15  
Q8QCN9 PRELIMINARY; PRT; 462 AA.  
ID Q8QCN9

J. Mol. Evol. 41:1048-1056(1995).  
EMBL; U19300; AAA97551.1; -  
DR HSSP; P46088; 2GSO.  
DR InterPro; IPR004046; GST\_Cterm.  
DR InterPro; IPR004045; GST\_Nterm.  
DR InterPro; IPR003083; S\_crystallin.  
DR Pfam; PF00043; GST\_C; 1.  
DR Pfam; PF02798; GST\_N; 1.  
DR PRINTS; PR01269; SCRSTALLIN.  
SQ SEQUENCE 217 AA; 26113 MW; 66D49D79D9B36F8A CRC64;

Query Match 57.4%; Score 39; DB 5; Length 217;  
Best Local Similarity 50.0%; Pred. No. 36;  
Matches 5; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 DVRYPTSRYLE 12  
|:|:|:|:|:  
Db 125 DLPYMQTRYME 136

RESULT 12  
Q8Q652 PRELIMINARY; PRT; 258 AA.  
ID Q8Q652 AC Q8Q652;  
DT 01-NOV-1998 (TrEMBLrel. 08, Created)  
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE T14N5.4 protein.  
GN T14N5.4.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Columbia;  
RA Federspiel N.A., Palm C.J., Conway A.B., Kurtz D.B., Conway A.R.,  
RA Au M., Araujo J., Buehler E., Dewar K., Feng J., Kim C., Li Y.,  
RA Oji O., Osborne B.I., Shinn P., Sun H., Toriumi M., Vysotskaia V.S.,  
RA Yu G., Ecker J., Theologis A., Davis R.W.;  
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AC004260; AAC34353.1; -  
DR F1R; T00447; T00447.  
DR InterPro; IPR006502; DUF\_plant\_1615.  
DR Pfam; PF04720; DUF506; 1.  
DR TIGRfams; TIGR01615; A\_thal\_3542; 1.  
SQ SEQUENCE 258 AA; 29578 MW; FBC341434DFA6380 CRC64;

Query Match 57.4%; Score 39; DB 10; Length 258;  
Best Local Similarity 66.7%; Pred. No. 43;  
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 4 PYRTSRYLE 12  
|:|:|:|:|:  
Db 199 PWRTSRMQ 207

RESULT 13  
Q88411 PRELIMINARY; PRT; 276 AA.  
ID Q88411 AC Q88411;  
DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE GTP cyclohydrolase I, putative.  
GN PSPT02112.  
OS Pseudomonas syringae (pv. tomato).  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
OC Pseudomonadaceae; Pseudomonas.  
OX NCBI\_TaxID=323;  
RN [1]  
RP SEQUENCE FROM N.A.

```

AC O8CON9;
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE Endo-alpha-D-mannosidase homolog.
GN 4932703L02RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs."
RL Nature 420:563-573 (2002).
DR EMBL; AK030141; BAC26805.1; -.
DR MGI; MGI:2444484; 4932703L02RIK.
SQ SEQUENCE 462 AA; 53199 MW; 5EA81B9C5075E684 CRC64;

Query Match          57.4%; Score 39; DB 11; Length 462;
Best Local Similarity 53.8%; Pred. No. 80;
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 DVPYRTSRYLEV 13
DB 426 DYRPHKPSLYLEL 438

```

Search completed: October 5, 2004, 08:13:11  
 Job time : 11.8812 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2004, 07:04:10 ; Search time 9.89043 Seconds  
(without alignments)  
371.381 Million cell updates/sec

Title: US-09-805-290A-8

Perfect score: 68

Sequence: 1 DVPFRTSRYLEV 13

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A Genesep 29Jan04:\*

1: genesep1980s:\*

2: genesep1990s:\*

3: genesep2000s:\*

4: genesep2001s:\*

5: genesep2002s:\*

6: genesep2003as:\*

7: genesep2003bs:\*

8: genesep2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	68	100.0	13	4	Aae10540 Llama spe
2	68	100.0	130	4	Aae10552 HPL inhib
3	67	98.5	13	4	Aae10542 Llama spe
4	67	98.5	130	4	Aae10559 HPL inhib
5	65	95.6	13	4	Aae10541 Llama spe
6	65	95.6	130	4	Aae10554 HPL inhib
7	64	94.1	13	4	Aae10543 Llama spe
8	42	61.8	727	6	Abu27962 Protein e
9	42	61.8	944	6	Abu45234 Protein e
10	42	61.8	961	6	Abu32202 Protein e
11	41	60.3	142	3	Abu45097 Human sec
12	41	60.3	451	7	Ade54766 Rat Prote
13	40	58.8	250	3	Aay73330 HTRM clon
14	40	58.8	252	3	Ag06656 Arabidops
15	40	58.8	272	3	Ag06655 Arabidops
16	40	58.8	281	3	Ag06654 Arabidops
17	39	57.4	208	3	Ag31529 Arabidops
18	39	57.4	238	3	Ag31528 Arabidops
19	39	57.4	258	3	Ag31527 Arabidops
20	39	57.4	509	5	Abu90986 Herbicida
21	39	57.4	558	6	Abu33563 Protein e
22	39	57.4	967	6	Abu99907 Human nov
23	39	57.4	1441	5	Aau75885 Human adh
24	38	55.9	52	4	Aau57763 Propionib
25	38	55.9	52	6	Abm54282 Propionib

26	38	55.9	90	6	ABP75655 Human sec
27	38	55.9	103	7	ADC89474 Ribosomal
28	38	55.9	254	4	ABP70887 Drosophil
29	38	55.9	262	6	ABU18705 Protein e
30	38	55.9	270	6	ABU17607 Protein e
31	38	55.9	465	6	ABU22918 Protein e
32	38	55.9	1718	5	Aae26420 Human tra
33	38	55.9	3503	4	ABP66499 Drosophil
34	37	54.4	58	2	Aaw87540 Peptide d
35	37	54.4	79	4	ABG24628 Novel hum
36	37	54.4	91	4	Aau62932 Propionib
37	37	54.4	91	6	ABM59451 Propionib
38	37	54.4	327	3	ABM18798 Partial a
39	37	54.4	391	7	ADC94988 E. faecul
40	37	54.4	437	7	ADC01045 Enterocae
41	37	54.4	566	2	Aaw87531 An antitu
42	37	54.4	600	6	ABM67347 Phototrab
43	37	54.4	614	4	ABG24803 Novel hum
44	37	54.4	620	5	ABP66018 Bifidobac
45	37	54.4	640	4	ABG03838 Novel hum

#### ALIGNMENTS

##### RESULT 1

AAE10540  
ID AAE10540 standard; peptide; 13 AA.

XX AAE10540;

DT 10-DEC-2001 (first entry)

DE Llama species antibody VHH CDR3 #9.

XX Llama antibody; camelid; anorectic; heavy chain variable domain; VHH;

KW human dietary enzyme inhibitor; medicament; human pancreatic lipase; HPL;

KW food; human Gastric lipase; HGL; cosmetic control; body weight;

KW complementarity determining region 3; CDR3.

XX Lama sp.

OS

PN EP1134231-A1.

XX 19-SEP-2001.

XX 20-FEB-2001; 2001EP-00200703.

XX 14-MAR-2000; 2000EP-00200930.

XX (UNIL ) UNILEVER NV.

XX (UNIL ) UNILEVER PLC.

XX Bezemer S, Van De Burg M, De Haard JJW, Tareilus E;

XX WPI; 2001-572718/65.

XX New antibody or its fragments for inhibiting human dietary enzymes,  
XX useful for cosmetic control of body weight of human beings, comprises  
XX heavy chain variable domain derived from immunoglobulin naturally devoid  
XX of light chains.

XX Disclosure; Page 17; 37pp; English.

XX The patent discloses antibodies or their fragments comprising a heavy  
XX chain variable domain (VHH) derived from an immunoglobulin naturally  
XX devoid of light chains specific for inhibiting human dietary enzymes. The  
XX antibodies of the invention are useful for the preparation of medicaments  
XX or food for inhibiting the activity of one or more human dietary enzymes  
XX especially human pancreatic lipase (HPL) or human gastric lipase (HGL)  
XX which are useful for the cosmetic control of body weight of human beings.  
XX The present peptide sequence is a complementarity determining region 3  
XX (CDR3) of llama species (camelid) antibody VHH region

XX SQ Sequence 13 AA;  
 Query Match 100.0%; Score 68; DB 4; Length 13;  
 Best Local Similarity 100.0%; Pred. No. 6.4e-06;  
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DVPRPRTSYLEV 13  
 |||||  
 Db 1 DVPRPRTSYLEV 13

RESULT 2  
 AAE10552  
 ID AAE10552 standard; peptide; 130 AA.  
 XX AC AAE10552;  
 XX DT 10-DEC-2001 (first entry)  
 XX DE HPL inhibiting VHH fragment, HPL #12 from llama species.  
 XX KW Llama antibody; camelid; anorectic; heavy chain variable domain; VHH;  
 XX KW human dietary enzyme inhibitor; medicament; human pancreatic lipase; HPL;  
 XX KW food; human gastric lipase; HGL; cosmetic control; body weight.  
 XX OS Lama sp.  
 XX FH Key  
 XX FT Region 31..35  
 XX FT /label= CDR1  
 XX FT /note= "Complementarity determining region 1"  
 XX FT Region 50..64  
 XX FT /label= CDR2  
 XX FT /note= "Complementarity determining region 2"  
 XX FT Region 98..110  
 XX FT /label= CDR3  
 XX FT /note= "Complementarity determining region 3"  
 XX PN EP1134231-A1.  
 XX PD 19-SEP-2001.  
 XX PF 20-FEB-2001; 2001EP-00200703.  
 XX PR 14-MAR-2000; 2000EP-00200930.  
 XX PA (UNIL ) UNILEVER NV.  
 XX PA (UNIL ) UNILEVER PLC.  
 XX PI Bezemer S, Van De Burg M, De Haard JJW, Tarsilus E;  
 XX WPI; 2001-572718/65.  
 XX PT New antibody or its fragments for inhibiting human dietary enzymes,  
 XX PT useful for cosmetic control of body weight of human beings, comprises  
 XX PT heavy chain variable domain derived from immunoglobulin naturally devoid  
 XX PT of light chains.  
 XX PS Example 2; Page 9; 37pp; English.  
 XX CC The patent discloses antibodies or their fragments comprising a heavy  
 XX CC chain variable domain (VHH) derived from an immunoglobulin naturally  
 XX CC devoid of light chains specific for inhibiting human dietary enzymes. The  
 XX CC antibodies of the invention are useful for the preparation of medicaments  
 XX CC or food for inhibiting the activity of one or more human dietary enzymes  
 XX CC especially human pancreatic lipase (HPL) or human gastric lipase (HGL)  
 XX CC which are useful for the cosmetic control of body weight of human beings.  
 XX CC The present peptide sequence is HPL inhibiting VHH fragment, HPL #12 from  
 XX CC llama (camelid) species  
 XX SQ Sequence 130 AA;

Query Match 100.0%; Score 68; DB 4; Length 130;  
 Best Local Similarity 100.0%; Pred. No. 8.3e-05;  
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DVPRPRTSYLEV 13  
 |||||  
 Db 98 DVPRPRTSYLEV 110

RESULT 3  
 AAE10542  
 ID AAE10542 standard; peptide; 13 AA.  
 XX AC AAE10542;  
 XX DT 10-DEC-2001 (first entry)  
 XX DE Llama species antibody VHH CDR3 #11.  
 XX KW Llama antibody; camelid; anorectic; heavy chain variable domain; VHH;  
 XX KW human dietary enzyme inhibitor; medicament; human pancreatic lipase; HPL;  
 XX KW food; human gastric lipase; HGL; cosmetic control; body weight;  
 XX KW complementarity determining region 3; CDR3.  
 XX OS Lama sp.  
 XX PN EP1134231-A1.  
 XX PD 19-SEP-2001.  
 XX PF 20-FEB-2001; 2001EP-00200703.  
 XX PR 14-MAR-2000; 2000EP-00200930.  
 XX PA (UNIL ) UNILEVER NV.  
 XX PA (UNIL ) UNILEVER PLC.  
 XX PI Bezemer S, Van De Burg M, De Haard JJW, Tarsilus E;  
 XX WPI; 2001-572718/65.  
 XX PT New antibody or its fragments for inhibiting human dietary enzymes,  
 XX PT useful for cosmetic control of body weight of human beings, comprises  
 XX PT heavy chain variable domain derived from immunoglobulin naturally devoid  
 XX PT of light chains.  
 XX PS Disclosure; Page 17; 37pp; English.  
 XX CC The patent discloses antibodies or their fragments comprising a heavy  
 XX CC chain variable domain (VHH) derived from an immunoglobulin naturally  
 XX CC devoid of light chains specific for inhibiting human dietary enzymes. The  
 XX CC antibodies of the invention are useful for the preparation of medicaments  
 XX CC or food for inhibiting the activity of one or more human dietary enzymes  
 XX CC especially human pancreatic lipase (HPL) or human gastric lipase (HGL)  
 XX CC which are useful for the cosmetic control of body weight of human beings.  
 XX CC The present peptide sequence is a complementarity determining region 3  
 XX CC (CDR3) of llama species (camelid) antibody VHH region  
 XX SQ Sequence 13 AA;

Query Match 98.5%; Score 67; DB 4; Length 13;  
 Best Local Similarity 92.3%; Pred. No. 9.9e-06;  
 Matches 12; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DVPRPRTSYLEV 13  
 |||||  
 Db 1 DVPRPRTSYLEV 13

RESULT 4  
 AAE10559  
 ID AAE10559 standard; peptide; 130 AA.  
 XX

```

AC AAE10559;
XX
XX 10-DEC-2001 (first entry)
XX
XX HPL inhibiting VHH fragment, HPL #30 from llama species.
XX
XX Llama antibody; camelid; anorectic; heavy chain variable domain; VHH;
XX human dietary enzyme inhibitor; medicament; human pancreatic lipase; HPL;
XX food; human gastric lipase; HGL; cosmetic control; body weight.
XX
XX Lama sp.
XX
XX Key Location/Qualifiers
XX Region 31..35
XX /label= CDR1
XX /note= "Complementarity determining region 1"
XX Region 49..64
XX /label= CDR2
XX /note= "Complementarity determining region 2"
XX Region 98..110
XX /label= CDR3
XX /note= "Complementarity determining region 3"
XX
XX EPI134231-A1.
XX
XX 19-SEP-2001.
XX
XX 20-FEB-2001; 2001EP-00200703.
XX
XX 14-MAR-2000; 2000EP-00200930.
XX
XX (UNIL ) UNILEVER NV.
XX (UNIL ) UNILEVER PLC.
XX
XX Bezemer S, Van De Burg M, De Haard JJW, Tareilus E;
XX WPI; 2001-572718/65.
XX
XX New antibody or its fragments for inhibiting human dietary enzymes,
XX useful for cosmetic control of body weight of human beings, comprises
XX heavy chain variable domain derived from immunoglobulin naturally devoid
XX of light chains.
XX
XX Example 2; Page 10; 37pp; English.
XX
XX The patent discloses antibodies or their fragments comprising a heavy
XX chain variable domain (VHH) derived from an immunoglobulin naturally
XX devoid of light chains specific for inhibiting human dietary enzymes. The
XX antibodies of the invention are useful for the preparation of medicaments
XX or food for inhibiting the activity of one or more human dietary enzymes
XX especially human pancreatic lipase (HPL) or human gastric lipase (HGL)
XX which are useful for the cosmetic control of body weight of human beings.
XX The present peptide sequence is HPL inhibiting VHH fragment, HPL #30 from
XX llama (camelid) species
XX
XX Sequence 130 AA;
XX
XX Query Match 98.5%; Score 67; DB 4; Length 130;
XX Best Local Similarity 92.3%; Pred. No. 0.00013;
XX Matches 12; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
XX
XX Qy 1 DVRPYRTSRYLEV 13
XX | | | | |
XX 98 DVRPYRTSRYLEI 110
XX
XX RESULT 5
XX AAE10541
XX ID AAE10541 standard; peptide; 13 AA.
XX
XX AC AAE10541;
XX
XX 10-DEC-2001 (first entry)
XX
XX

```

```

XX
XX Llama species antibody VHH CDR3 #10.
XX
XX Llama antibody; camelid; anorectic; heavy chain variable domain; VHH;
XX human dietary enzyme inhibitor; medicament; human pancreatic lipase; HPL;
XX food; human gastric lipase; HGL; cosmetic control; body weight;
XX complementarity determining region 3; CDR3.
XX
XX Lama sp.
XX
XX EPI134231-A1.
XX
XX 19-SEP-2001.
XX
XX 20-FEB-2001; 2001EP-00200703.
XX
XX 14-MAR-2000; 2000EP-00200930.
XX
XX (UNIL ) UNILEVER NV.
XX (UNIL ) UNILEVER PLC.
XX
XX Bezemer S, Van De Burg M, De Haard JJW, Tareilus E;
XX WPI; 2001-572718/65.
XX
XX New antibody or its fragments for inhibiting human dietary enzymes,
XX useful for cosmetic control of body weight of human beings, comprises
XX heavy chain variable domain derived from immunoglobulin naturally devoid
XX of light chains.
XX
XX Disclosure; Page 17; 37pp; English.
XX
XX The patent discloses antibodies or their fragments comprising a heavy
XX chain variable domain (VHH) derived from an immunoglobulin naturally
XX devoid of light chains specific for inhibiting human dietary enzymes. The
XX antibodies of the invention are useful for the preparation of medicaments
XX or food for inhibiting the activity of one or more human dietary enzymes
XX especially human pancreatic lipase (HPL) or human gastric lipase (HGL)
XX which are useful for the cosmetic control of body weight of human beings.
XX The present peptide sequence is a complementarity determining region 3
XX (CDR3) of llama species (camelid) antibody VHH region
XX
XX Sequence 13 AA;
XX
XX Query Match 95.6%; Score 65; DB 4; Length 13;
XX Best Local Similarity 92.3%; Pred. No. 2.4e-05;
XX Matches 12; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
XX
XX Qy 1 DVRPYRTSRYLEV 13
XX | | | | |
XX Db 1 DVRPYRTSRYLEL 13
XX
XX RESULT 6
XX AAE10554
XX ID AAE10554 standard; peptide; 130 AA.
XX
XX AC AAE10554;
XX
XX 10-DEC-2001 (first entry)
XX
XX HPL inhibiting VHH fragment, HPL #14 from llama species.
XX
XX Llama antibody; camelid; anorectic; heavy chain variable domain; VHH;
XX human dietary enzyme inhibitor; medicament; human pancreatic lipase; HPL;
XX food; human gastric lipase; HGL; cosmetic control; body weight.
XX
XX Lama sp.
XX
XX Key Location/Qualifiers
XX Region 31..35
XX /label= CDR1
XX /note= "Complementarity determining region 1"
XX
XX

```



DR N-PSDB; ACA31832.  
 XX New antisense nucleic acids, useful for identifying proteins or screening  
 PT for homologous nucleic acids required for cellular proliferation to  
 PT isolate candidate molecules for rational drug discovery programs.  
 XX  
 XX Claim 25; SEQ ID NO 55886; 1766pp; English.  
 XX  
 XX The invention relates to an isolated nucleic acid comprising any one of  
 CC the 6213 antisense sequences given in the specification where expression  
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
 CC (1) a vector comprising a promoter operably linked to the nucleic acid  
 CC encoding a polypeptide whose expression is inhibited by the antisense  
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
 CC polypeptide or its fragment whose expression is inhibited by the  
 CC antisense nucleic acid; (4) an antibody capable of specifically binding  
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
 CC proliferation or the activity of a gene in an operon required for  
 CC proliferation; (7) identifying a compound that influences the activity of  
 CC the gene product or that has an activity against a biological pathway  
 CC required for proliferation, or that inhibits cellular proliferation; (8)  
 CC identifying a gene required for cellular proliferation or the biological  
 CC pathway in which a proliferation-required gene or its gene product lies  
 CC or a gene on which the test compound that inhibits proliferation of an  
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
 CC compound's activity; (11) a culture comprising strains in which the gene  
 CC product is overexpressed or underexpressed; (12) determining the extent  
 CC to which each of the strains is present in a culture or collection of  
 CC strains; or (13) identifying the target of a compound that inhibits the  
 CC proliferation of an organism. The antisense nucleic acids are useful for  
 CC identifying proteins or screening for homologous nucleic acids required  
 CC for cellular proliferation to isolate candidate molecules for rational  
 CC drug discovery programs, or for screening homologous nucleic acids  
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
 CC the target prokaryotic essential genes. Note: The sequence data for this  
 CC patent did not form part of the printed specification, but was obtained  
 CC in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 XX SQ Sequence 727 AA;  
 Query Match 61.8%; Score 42; DB 6; Length 727;  
 Best Local Similarity 80.0%; Pred. No. 47;  
 Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 QY 4 PYRTSRYLEV 13  
 ||||: ||||  
 Db 119 PYRTAFYLEV 128  
 RESULT 9  
 ABU45234  
 ID ABU45234 standard; protein; 944 AA.  
 XX  
 XX AC ABU45234;  
 XX  
 XX DT 19-JUN-2003 (first entry)  
 XX  
 XX DE Protein encoded by Prokaryotic essential gene #30761.  
 XX  
 XX KW Antisense; prokaryotic essential gene; cell proliferation; drug design.  
 XX OS Salmonella paratyphi.  
 XX  
 XX PN WO20027183-A2.  
 XX  
 XX PD 03-OCT-2002.  
 XX  
 XX PF 21-MAR-2002; 2002WO-US009107.  
 XX  
 XX PR 21-MAR-2001; 2001US-00815242.  
 PR 06-SEP-2001; 2001US-00948993.

PR 25-OCT-2001; 2001US-0342923P.  
 PR 08-FEB-2002; 2002US-00072851.  
 PR 06-MAR-2002; 2002US-0362699P.  
 XX  
 XX (ELIT-) ELITRA PHARM INC.  
 XX  
 XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
 PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
 XX  
 XX WPI; 2003-029926/02.  
 DR N-PSDB; ACA49104.  
 XX  
 XX New antisense nucleic acids, useful for identifying proteins or screening  
 PT for homologous nucleic acids required for cellular proliferation to  
 PT isolate candidate molecules for rational drug discovery programs.  
 XX  
 XX Claim 25; SEQ ID NO 73158; 1766pp; English.  
 XX  
 XX The invention relates to an isolated nucleic acid comprising any one of  
 CC the 6213 antisense sequences given in the specification where expression  
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
 CC (1) a vector comprising a promoter operably linked to the nucleic acid  
 CC encoding a polypeptide whose expression is inhibited by the antisense  
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
 CC polypeptide or its fragment whose expression is inhibited by the  
 CC antisense nucleic acid; (4) an antibody capable of specifically binding  
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
 CC proliferation or the activity of a gene in an operon required for  
 CC proliferation; (7) identifying a compound that influences the activity of  
 CC the gene product or that has an activity against a biological pathway  
 CC required for proliferation, or that inhibits cellular proliferation; (8)  
 CC identifying a gene required for cellular proliferation or the biological  
 CC pathway in which a proliferation-required gene or its gene product lies  
 CC or a gene on which the test compound that inhibits proliferation of an  
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
 CC compound's activity; (11) a culture comprising strains in which the gene  
 CC product is overexpressed or underexpressed; (12) determining the extent  
 CC to which each of the strains is present in a culture or collection of  
 CC strains; or (13) identifying the target of a compound that inhibits the  
 CC proliferation of an organism. The antisense nucleic acids are useful for  
 CC identifying proteins or screening for homologous nucleic acids required  
 CC for cellular proliferation to isolate candidate molecules for rational  
 CC drug discovery programs, or for screening homologous nucleic acids  
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
 CC the target prokaryotic essential genes. Note: The sequence data for this  
 CC patent did not form part of the printed specification, but was obtained  
 CC in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 XX SQ Sequence 944 AA;  
 Query Match 61.8%; Score 42; DB 6; Length 944;  
 Best Local Similarity 80.0%; Pred. No. 63;  
 Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 QY 4 PYRTSRYLEV 13  
 ||||: ||||  
 Db 115 PYRTAFYLEV 124  
 RESULT 10  
 ABU32202  
 ID ABU32202 standard; protein; 961 AA.  
 XX  
 XX AC ABU32202;  
 XX  
 XX DT 19-JUN-2003 (first entry)  
 XX  
 XX DE Protein encoded by Prokaryotic essential gene #17729.  
 XX  
 XX KW Antisense; prokaryotic essential gene; cell proliferation; drug design.  
 XX





QY 1 DVPYRTSRYLEV 13  
Db 415 DYRPHKPSLYLEI 427

RESULT 13  
AAAY73330  
ID AAY73330 standard; protein; 250 AA.  
XX  
AC AAY73330;  
XX  
DT 24-FEB-2000 (first entry)  
XX  
DE HTRM clone 397663 protein sequence.  
XX  
KW HTRM; human transcriptional regulatory molecule; arteriosclerosis; AIDS;  
KW arteriosclerosis; cirrhosis; cancer; leukaemia; diabetes mellitus;  
KW Addison's disease; multiple sclerosis; rheumatoid arthritis; infection;  
KW trauma; myasthenia gravis; adenocarcinoma; immune disorder; treatment.  
XX  
OS Homo sapiens.  
XX  
PN WO9957144-A2.  
XX  
PD 11-NOV-1999.  
XX  
PF 04-MAY-1999; 99WO-US009935.  
XX  
PR 05-MAY-1998; 98US-0084254P.  
PR 07-AUG-1998; 98US-0095827P.  
PR 02-OCT-1998; 98US-0102745P.  
XX  
PA (INCY-) INCYTE PHARM INC.  
XX  
PI Hillman JL, Bandman O, Lal P, Yue H, Reddy R, Tang YT;  
PI Gerstein EH, Patterson C, Baughn MR, Azimzai Y, Lu DAM;  
XX  
DR WPI; 2000-052941/04.  
DR N-PSDB; AA52415.  
XX  
PT New peptides useful for diagnosis, prevention and treatment of cancer and  
PT immune disorders.  
XX  
PS Claim 1; Page 92-93; 193pp; English.  
XX  
CC AAY73325-Y73389 are human transcriptional regulator molecule (HTRM)  
CC protein sequences. The HTRM protein and nucleotide sequences are useful  
CC for preventing or treating disorders associated with decreased expression  
CC or activity of HTRM which include cell proliferative disorders such as  
CC arteriosclerosis and cirrhosis; cancers including adenocarcinoma and  
CC leukaemia; immune disorders such as AIDS, Addison's disease, diabetes  
CC mellitus, rheumatoid arthritis, multiple sclerosis, systemic lupus  
CC erythematosus, and myasthenia gravis; infections and trauma. Antagonists  
CC of the HTRM polypeptides are useful for treating or preventing disorders  
CC associated with increased expression or activity of HTRMs. HTRM  
CC polypeptides, their immunogenic fragments or oligopeptides are useful for  
CC screening libraries of compounds in drug screening techniques.  
CC Polynucleotides encoding HTRM are useful for blocking the transcription  
CC of mRNA and regulating gene function by modulating the activity of HTRM.  
CC Vectors expressing HTRM or agonists can also be used to prevent or treat  
CC disorder associated with decreased HTRM expression. Antibodies which  
CC specifically bind HTRM and polynucleotides encoding HTRM are useful for  
CC diagnosing disorders associated with the expression of HTRM, particularly  
CC in assays that detect the expression of HTRM. Nucleotide sequences  
CC encoding HTRM may be useful to generate hybridization probes useful in  
CC mapping the naturally occurring genomic sequence and to detect  
CC differences in gene sequences among normal, carrier and affected  
CC individuals. Using diagnostic assays, cancer can be detected prior to the  
CC appearance of clinical symptoms and thereby progression of cancer can be  
CC prevented by aggressive treatment or preventive measures  
XX  
SQ Sequence 250 AA;

QY 1 DVPYRTSRYLEV 13  
Db 415 DYRPHKPSLYLEI 427

RESULT 12  
ADE54766  
ID ADE54766 standard; protein; 451 AA.  
XX  
AC ADE54766;  
XX  
DT 29-JAN-2004 (first entry)  
XX  
DE Rat Protein AAB86925, SEQ ID NO 571.  
XX  
KW Rat; pain; neuronal tissue; gene therapy; spinal segmental nerve injury;  
KW chronic constriction injury; CCI; spared nerve injury; SNI; Chung.  
XX  
OS Rattus norvegicus.  
XX  
PN WO2003016475-A2.  
XX  
PD 27-FEB-2003.  
XX  
PF 14-AUG-2002; 2002WO-US025765.  
XX  
PR 14-AUG-2001; 2001US-0312147P.  
PR 01-NOV-2001; 2001US-0346382P.  
PR 26-NOV-2001; 2001US-0333347P.  
XX  
PA (GEO) GEN HOSPITAL CORP.  
PA (FARB) BAYER AG.  
XX  
PI Woolf C, D'urso D, Befort K, Costigan M;  
XX  
DR WPI; 2003-268312/26.  
DR GENBANK; AAB86925.  
XX  
PT New composition comprising two or more isolated polypeptides, useful for  
PT preparing a medicament for treating pain in an animal.  
XX  
PS Claim 1; Page; 1017pp; English.  
XX  
CC The invention discloses a composition comprising two or more isolated rat  
CC or human polynucleotides or a polynucleotide which represents a fragment,  
CC derivative or allelic variation of the nucleic acid sequence. Also  
CC claimed are a vector comprising the novel polynucleotide, a host cell  
CC comprising the vector, a method for identifying a nucleotide sequence  
CC which is differentially regulated in an animal subjected to pain and a  
CC kit to perform the method, an array, a method for identifying an agent  
CC that increases or decreases the expression of the polynucleotide sequence  
CC that is differentially expressed in neuronal tissue of a first animal  
CC subjected to pain, a method for identifying a compound which regulates  
CC the expression of a polynucleotide sequence which is differentially  
CC expressed in an animal subjected to pain, a method for identifying a  
CC compound that regulates the activity of one or more of the  
CC polynucleotides, a method for producing a pharmaceutical composition, a  
CC method for identifying a compound or small molecule that regulates the  
CC activity in an animal of one or more of the polypeptides given in the  
CC specification, a method for identifying a compound useful in treating  
CC pain and a pharmaceutical composition comprising the one or more  
CC polypeptides or their antibodies. The polynucleotide or the compound that  
CC modulates its activity is useful for preparing a medicament for treating  
CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction  
CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene  
CC therapy). The sequence presented is a rat protein (shown in Table 2 of  
CC the specification) which is differentially expressed during pain. Note:  
CC The sequence data for this patent did not form part of the printed  
CC ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 451 AA;

Query Match 60.3%; Score 41; DB 7; Length 451;  
Best Local Similarity 53.8%; Pred. No. 43;  
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Query Match 58.8%; Score 40; DB 3; Length 250;  
 Best Local Similarity 66.7%; Pred. No. 34;  
 Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 VRPYRTSRVLEV 13  
 Db 44 VRPHRSRQLQV 55

## RESULT 14

AAG06656  
 ID AAG06656 standard; protein; 252 AA.

XX AC AAG06656;

XX DT 17-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 3507.

XX KW Protein identification; signal transduction pathway; metabolic pathway;  
 KW hybridisation assay; genetic mapping; gene expression control; promoter;  
 KW termination sequence.

XX OS Arabidopsis thaliana.

XX PN EF1033405-A2.

XX PD 06-SEP-2000.

XX PF 25-FEB-2000; 2000EP-00301439.

XX PR 25-FEB-1999; 99US-0121825P.

PR 05-MAR-1999; 99US-0123180P.

PR 09-MAR-1999; 99US-0123548P.

PR 23-MAR-1999; 99US-0125788P.

PR 25-MAR-1999; 99US-0126264P.

PR 29-MAR-1999; 99US-0126785P.

PR 01-APR-1999; 99US-0127462P.

PR 06-APR-1999; 99US-0128234P.

PR 08-APR-1999; 99US-0128714P.

PR 16-APR-1999; 99US-0129645P.

PR 19-APR-1999; 99US-0130077P.

PR 21-APR-1999; 99US-0130449P.

PR 23-APR-1999; 99US-0130510P.

PR 23-APR-1999; 99US-0130891P.

PR 28-APR-1999; 99US-0134499P.

PR 30-APR-1999; 99US-0132048P.

PR 30-APR-1999; 99US-0132407P.

PR 04-MAY-1999; 99US-0132484P.

PR 05-MAY-1999; 99US-0132485P.

PR 06-MAY-1999; 99US-0132486P.

PR 06-MAY-1999; 99US-0132487P.

PR 07-MAY-1999; 99US-01324863P.

PR 11-MAY-1999; 99US-0132456P.

PR 14-MAY-1999; 99US-0134218P.

PR 14-MAY-1999; 99US-0134219P.

PR 14-MAY-1999; 99US-0134221P.

PR 14-MAY-1999; 99US-0134370P.

PR 16-MAY-1999; 99US-0134768P.

PR 19-MAY-1999; 99US-0134941P.

PR 20-MAY-1999; 99US-0135124P.

PR 21-MAY-1999; 99US-0135353P.

PR 24-MAY-1999; 99US-0135629P.

PR 25-MAY-1999; 99US-0136021P.

PR 27-MAY-1999; 99US-0136392P.

PR 28-MAY-1999; 99US-0136782P.

PR 01-JUN-1999; 99US-0137222P.

PR 03-JUN-1999; 99US-0137528P.

PR 04-JUN-1999; 99US-0137502P.

PR 07-JUN-1999; 99US-0137724P.

PR 08-JUN-1999; 99US-0138094P.

PR 10-JUN-1999; 99US-0138540P.

PR 10-JUN-1999; 99US-0138847P.  
 PR 14-JUN-1999; 99US-0139119P.  
 PR 16-JUN-1999; 99US-0139452P.  
 PR 17-JUN-1999; 99US-0139453P.  
 PR 18-JUN-1999; 99US-0139452P.  
 PR 18-JUN-1999; 99US-0139454P.  
 PR 18-JUN-1999; 99US-0139455P.  
 PR 18-JUN-1999; 99US-0139456P.  
 PR 18-JUN-1999; 99US-0139457P.  
 PR 18-JUN-1999; 99US-0139458P.  
 PR 18-JUN-1999; 99US-0139459P.  
 PR 18-JUN-1999; 99US-0139460P.  
 PR 18-JUN-1999; 99US-0139461P.  
 PR 18-JUN-1999; 99US-0139462P.  
 PR 18-JUN-1999; 99US-0139463P.  
 PR 18-JUN-1999; 99US-0139750P.  
 PR 18-JUN-1999; 99US-0139763P.  
 PR 21-JUN-1999; 99US-0139817P.  
 PR 22-JUN-1999; 99US-0139899P.  
 PR 23-JUN-1999; 99US-0140353P.  
 PR 23-JUN-1999; 99US-0140354P.  
 PR 24-JUN-1999; 99US-0140695P.  
 PR 28-JUN-1999; 99US-0140823P.  
 PR 29-JUN-1999; 99US-0140921P.  
 PR 30-JUN-1999; 99US-0141287P.  
 PR 01-JUL-1999; 99US-0141842P.  
 PR 01-JUL-1999; 99US-0142154P.  
 PR 02-JUL-1999; 99US-0142055P.  
 PR 06-JUL-1999; 99US-0142390P.  
 PR 08-JUL-1999; 99US-0142803P.  
 PR 09-JUL-1999; 99US-0142920P.  
 PR 12-JUL-1999; 99US-0142977P.  
 PR 13-JUL-1999; 99US-0143542P.  
 PR 14-JUL-1999; 99US-0143624P.  
 PR 15-JUL-1999; 99US-0144005P.  
 PR 16-JUL-1999; 99US-0144085P.  
 PR 16-JUL-1999; 99US-0144086P.  
 PR 19-JUL-1999; 99US-0144325P.  
 PR 19-JUL-1999; 99US-0144331P.  
 PR 19-JUL-1999; 99US-0144332P.  
 PR 19-JUL-1999; 99US-0144333P.  
 PR 19-JUL-1999; 99US-0144334P.  
 PR 19-JUL-1999; 99US-0144335P.  
 PR 20-JUL-1999; 99US-0144352P.  
 PR 20-JUL-1999; 99US-0144632P.  
 PR 20-JUL-1999; 99US-0144884P.  
 PR 20-JUL-1999; 99US-0144814P.  
 PR 21-JUL-1999; 99US-0145086P.  
 PR 21-JUL-1999; 99US-0145088P.  
 PR 22-JUL-1999; 99US-0145085P.  
 PR 22-JUL-1999; 99US-0145087P.  
 PR 22-JUL-1999; 99US-0145089P.  
 PR 22-JUL-1999; 99US-0145192P.  
 PR 23-JUL-1999; 99US-0145145P.  
 PR 23-JUL-1999; 99US-0145218P.  
 PR 23-JUL-1999; 99US-0145224P.  
 PR 26-JUL-1999; 99US-0145276P.  
 PR 27-JUL-1999; 99US-0145913P.  
 PR 27-JUL-1999; 99US-0145918P.  
 PR 27-JUL-1999; 99US-0145919P.  
 PR 28-JUL-1999; 99US-0145951P.  
 PR 02-AUG-1999; 99US-0146386P.  
 PR 02-AUG-1999; 99US-0146388P.  
 PR 02-AUG-1999; 99US-0146389P.  
 PR 03-AUG-1999; 99US-0147038P.  
 PR 04-AUG-1999; 99US-0147204P.  
 PR 04-AUG-1999; 99US-0147302P.  
 PR 05-AUG-1999; 99US-0147192P.  
 PR 05-AUG-1999; 99US-0147260P.  
 PR 06-AUG-1999; 99US-0147303P.  
 PR 06-AUG-1999; 99US-0147416P.  
 PR 08-AUG-1999; 99US-0147493P.  
 PR 09-AUG-1999; 99US-0147935P.

	77	DTHPRMSRYIQL	89
Db			
		RESULT.15	
	AAG06655		
ID	AAG06655 standard; protein; 272 AA.		
XX			
AC	AAG06655;		
XX			
DT	17-OCT-2000 (first entry)		
DE	Arabidopsis thaliana protein fragment SEQ ID NO: 3506.		
XX			
KW	Protein identification; signal transduction pathway; metabolic pathway;		
KW	hybridisation assay; generic mapping; gene expression control; promoter;		
KW	termination sequence.		
XX			
OS	Arabidopsis thaliana.		
XX			
PX	EP1033405-A2.		
XX			
PD	06-SEP-2000.		
XX			
PF	25-FEB-2000; 2000EP-00301439.		
XX			
PR	25-FEB-1999; 99US-0121825P.		
PR	05-MAR-1999; 99US-0123180P.		
PR	09-MAR-1999; 99US-0123548P.		
PR	23-VAR-1999; 99US-0125788P.		
PR	25-MAR-1999; 99US-0126264P.		
PR	29-NAR-1999; 99US-0126785P.		
PR	01-APR-1999; 99US-0127462P.		
PR	06-APR-1999; 99US-0128234P.		
PR	08-APR-1999; 99US-0128714P.		
PR	16-APR-1999; 99US-0129845P.		
PR	19-APR-1999; 99US-0130077P.		
PR	21-APR-1999; 99US-0130449P.		
PR	23-APR-1999; 99US-0130510P.		
PR	23-APR-1999; 99US-0130891P.		
PR	28-APR-1999; 99US-0131449P.		
PR	30-APR-1999; 99US-0132048P.		
PR	04-MAY-1999; 99US-0132407P.		
PR	05-MAY-1999; 99US-0132484P.		
PR	06-MAY-1999; 99US-0132485P.		
PR	07-MAY-1999; 99US-0132487P.		
PR	11-MAY-1999; 99US-0132863P.		
PR	14-MAY-1999; 99US-0134256P.		
PR	14-MAY-1999; 99US-0134218P.		
PR	14-MAY-1999; 99US-0134219P.		
PR	14-MAY-1999; 99US-0134221P.		
PR	14-MAY-1999; 99US-0134370P.		
PR	18-MAY-1999; 99US-0134768P.		
PR	19-MAY-1999; 99US-0134941P.		
PR	20-MAY-1999; 99US-0135124P.		
PR	21-MAY-1999; 99US-0135353P.		
PR	24-MAY-1999; 99US-0135629P.		
PR	25-MAY-1999; 99US-0136021P.		
PR	27-MAY-1999; 99US-0136392P.		
PR	28-MAY-1999; 99US-0136782P.		
PR	01-JUN-1999; 99US-0137222P.		
PR	04-JUN-1999; 99US-0137528P.		
PR	07-JUN-1999; 99US-0137502P.		
PR	08-JUN-1999; 99US-0137724P.		
PR	10-JUN-1999; 99US-0138094P.		
PR	10-JUN-1999; 99US-0138540P.		
PR	14-JUN-1999; 99US-0138847P.		
PR	14-JUN-1999; 99US-0139119P.		
PR	16-JUN-1999; 99US-0139452P.		
PR	16-JUN-1999; 99US-0139453P.		
PR	17-JUN-1999; 99US-0139492P.		
PR	18-JUN-1999; 99US-0139454P.		
PR	18-JUN-1999; 99US-0139455P.		

Query Match      58.8; Score 40; DB 3; Length 252;  
Best Local Similarity    46.2; Pred.No. 34;  
Matches     6; Conservative       4; Mismatches                  3; Indels                  0; Gaps                  0;

QY  
  
1 DVRPYRTSRYLEV 13  
|...||..|  
+-----+

PR 18-JUN-1999; 99US-0139456P.  
 PR 18-JUN-1999; 99US-0139457P.  
 PR 18-JUN-1999; 99US-0139458P.  
 PR 18-JUN-1999; 99US-0139459P.  
 PR 18-JUN-1999; 99US-0139460P.  
 PR 18-JUN-1999; 99US-0139461P.  
 PR 18-JUN-1999; 99US-0139462P.  
 PR 18-JUN-1999; 99US-0139463P.  
 PR 18-JUN-1999; 99US-0139750P.  
 PR 18-JUN-1999; 99US-0139763P.  
 PR 21-JUN-1999; 99US-0139817P.  
 PR 22-JUN-1999; 99US-0139899P.  
 PR 23-JUN-1999; 99US-0140353P.  
 PR 23-JUN-1999; 99US-0140354P.  
 PR 24-JUN-1999; 99US-0140355P.  
 PR 28-JUN-1999; 99US-0140823P.  
 PR 28-JUN-1999; 99US-0140921P.  
 PR 30-JUN-1999; 99US-0141287P.  
 PR 01-JUL-1999; 99US-0141842P.  
 PR 02-JUL-1999; 99US-0142154P.  
 PR 06-JUL-1999; 99US-0142055P.  
 PR 08-JUL-1999; 99US-0142390P.  
 PR 08-JUL-1999; 99US-0142803P.  
 PR 08-JUL-1999; 99US-0142820P.  
 PR 12-JUL-1999; 99US-0142977P.  
 PR 13-JUL-1999; 99US-0143354P.  
 PR 14-JUL-1999; 99US-0143624P.  
 PR 15-JUL-1999; 99US-0144005P.  
 PR 16-JUL-1999; 99US-0144085P.  
 PR 18-JUL-1999; 99US-0144086P.  
 PR 18-JUL-1999; 99US-0144325P.  
 PR 19-JUL-1999; 99US-0144331P.  
 PR 19-JUL-1999; 99US-0144332P.  
 PR 19-JUL-1999; 99US-0144333P.  
 PR 19-JUL-1999; 99US-0144334P.  
 PR 19-JUL-1999; 99US-0144335P.  
 PR 20-JUL-1999; 99US-0144352P.  
 PR 20-JUL-1999; 99US-0144632P.  
 PR 20-JUL-1999; 99US-0144884P.  
 PR 21-JUL-1999; 99US-0144814P.  
 PR 21-JUL-1999; 99US-0145086P.  
 PR 21-JUL-1999; 99US-0145088P.  
 PR 22-JUL-1999; 99US-0145085P.  
 PR 22-JUL-1999; 99US-0145087P.  
 PR 22-JUL-1999; 99US-0145089P.  
 PR 23-JUL-1999; 99US-0145192P.  
 PR 23-JUL-1999; 99US-0145145P.  
 PR 23-JUL-1999; 99US-0145218P.  
 PR 26-JUL-1999; 99US-0145224P.  
 PR 27-JUL-1999; 99US-0145276P.  
 PR 27-JUL-1999; 99US-0145913P.  
 PR 27-JUL-1999; 99US-0145918P.  
 PR 27-JUL-1999; 99US-0145919P.  
 PR 28-JUL-1999; 99US-0145951P.  
 PR 02-AUG-1999; 99US-0146386P.  
 PR 02-AUG-1999; 99US-0146388P.  
 PR 02-AUG-1999; 99US-0146389P.  
 PR 03-AUG-1999; 99US-0147038P.  
 PR 04-AUG-1999; 99US-0147204P.  
 PR 04-AUG-1999; 99US-0147302P.  
 PR 05-AUG-1999; 99US-0147192P.  
 PR 05-AUG-1999; 99US-0147260P.  
 PR 06-AUG-1999; 99US-0147303P.  
 PR 06-AUG-1999; 99US-0147416P.  
 PR 09-AUG-1999; 99US-0147493P.  
 PR 09-AUG-1999; 99US-0147935P.  
 PR 10-AUG-1999; 99US-0148171P.  
 PR 11-AUG-1999; 99US-0148319P.  
 PR 12-AUG-1999; 99US-0148341P.  
 PR 13-AUG-1999; 99US-0148565P.  
 PR 13-AUG-1999; 99US-0148684P.  
 PR 16-AUG-1999; 99US-0149368P.  
 PR 17-AUG-1999; 99US-0149175P.

PR 18-AUG-1999; 99US-0149426P.  
 PR 20-AUG-1999; 99US-0149722P.  
 PR 20-AUG-1999; 99US-0149723P.  
 PR 20-AUG-1999; 99US-0149929P.  
 PR 23-AUG-1999; 99US-0149902P.  
 PR 23-AUG-1999; 99US-0149930P.  
 PR 25-AUG-1999; 99US-0150566P.  
 PR 26-AUG-1999; 99US-0150844P.  
 PR 27-AUG-1999; 99US-0151085P.  
 PR 27-AUG-1999; 99US-0151086P.  
 PR 27-AUG-1999; 99US-0151080P.  
 PR 30-AUG-1999; 99US-0151303P.  
 PR 31-AUG-1999; 99US-0151438P.  
 PR 01-SEP-1999; 99US-0151930P.  
 PR 07-SEP-1999; 99US-0152363P.  
 PR 10-SEP-1999; 99US-0153070P.  
 PR 13-SEP-1999; 99US-0153758P.  
 PR 15-SEP-1999; 99US-0154018P.  
 PR 16-SEP-1999; 99US-0154039P.  
 PR 20-SEP-1999; 99US-0154779P.  
 PR 22-SEP-1999; 99US-0155139P.  
 PR 24-SEP-1999; 99US-0155486P.  
 PR 24-SEP-1999; 99US-0155859P.  
 PR 28-SEP-1999; 99US-0156458P.  
 PR 29-SEP-1999; 99US-0156596P.  
 PR 05-OCT-1999; 99US-0157117P.  
 PR 06-OCT-1999; 99US-0157865P.  
 PR 07-OCT-1999; 99US-0158029P.  
 PR 08-OCT-1999; 99US-0158232P.  
 PR 12-OCT-1999; 99US-0158369P.  
 PR 13-OCT-1999; 99US-0159293P.  
 PR 13-OCT-1999; 99US-0159294P.  
 PR 13-OCT-1999; 99US-0159295P.  
 PR 14-OCT-1999; 99US-0159329P.  
 PR 14-OCT-1999; 99US-0159330P.  
 PR 14-OCT-1999; 99US-0159331P.  
 PR 14-OCT-1999; 99US-0159637P.  
 PR 14-OCT-1999; 99US-0159638P.  
 PR 18-OCT-1999; 99US-0159584P.  
 PR 21-OCT-1999; 99US-0160741P.  
 PR 21-OCT-1999; 99US-0160767P.  
 PR 21-OCT-1999; 99US-0160768P.  
 PR 21-OCT-1999; 99US-0160770P.  
 PR 21-OCT-1999; 99US-0160814P.  
 PR 21-OCT-1999; 99US-0160815P.  
 PR 22-OCT-1999; 99US-0160980P.  
 PR 22-OCT-1999; 99US-0160981P.  
 PR 22-OCT-1999; 99US-0160989P.  
 PR 25-OCT-1999; 99US-0161404P.  
 PR 25-OCT-1999; 99US-0161405P.  
 PR 25-OCT-1999; 99US-0161406P.  
 PR 26-OCT-1999; 99US-0161359P.  
 PR 26-OCT-1999; 99US-0161360P.  
 PR 26-OCT-1999; 99US-0161361P.  
 PR 28-OCT-1999; 99US-0161920P.  
 PR 28-OCT-1999; 99US-0161922P.  
 PR 28-OCT-1999; 99US-0161993P.  
 PR 29-OCT-1999; 99US-0162142P.

Query Match 58.8%; Score 40; DB 3; Length 272;  
 Best Local Similarity 46.2%; Pred. No. 38;  
 Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Qy 1 DVPPTTSRYLEV 13  
 Db 97 DTHPHMSRYIOL 109

Search completed: October 5, 2004, 07:59:36  
 Job time : 13.8904 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2004, 07:36:16 ; Search time 2.10648 Seconds  
(without alignments)  
593.639 Million cell updates/sec

Title: US-09-805-290A-9

Perfect score: 68

Sequence: 1 DVRPYRTSRYLEL 13

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: Pir.78.\*

2: Pir1.\*

3: Pir2.\*

4: Pir3.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	41	60.3	775	2 B84828	probable WD-40 rep
2	40	58.8	437	2 G91067	hypothetical prote
3	40	58.8	441	2 H85911	hypothetical prote
4	40	58.8	442	2 A3042	oxidoreductase Atu
5	40	58.8	442	2 H88243	agaE protein [impo
6	39	57.4	258	2 T00447	hypothetical prote
7	39	57.4	303	2 T47356	hypothetical prote
8	39	57.4	526	2 A86274	F7A19.15 protein -
9	39	57.4	569	2 B82559	30S ribosomal prot
10	39	57.4	962	1 SNECPI	pitriylisin (EC 3.4
11	39	57.4	962	2 F91088	proteinase III [im
12	39	57.4	962	2 H85933	proteinase III [im
13	39	57.4	962	2 AC0865	protease III precu
14	38	55.9	103	2 T24300	ribosomal protein
15	38	55.9	2870	2 H36974	cyclic beta 1-2 gl
16	37	54.4	64	2 A81857	hypothetical DNA-b
17	37	54.4	286	2 T08250	hypothetical prote
18	37	54.4	594	2 T15202	hypothetical prote
19	37	54.4	599	2 T10798	phosphorin-S - vo
20	36.5	53.7	496	2 S75790	hypothetical prote
21	36	52.9	168	2 H72721	hypothetical prote
22	36	52.9	260	2 T00446	hypothetical prote
23	36	52.9	396	2 H83576	methionine adenosy
24	36	52.9	415	2 G30705	probable transport
25	36	52.9	415	2 B85556	probable transport
26	36	52.9	415	2 G84790	yodG protein - Esc
27	36	52.9	415	2 AB0573	probable membrane
28	36	52.9	432	1 JN0317	protein-tyrosine-p
29	36	52.9	432	1 A34845	protein-tyrosine-p

30 36 52.9 435 2 A72658 probable isocitrat  
31 36 52.9 536 2 E84465 hypothetical prote  
32 36 52.9 556 2 B82141 ribosomal protein  
33 36 52.9 559 2 C83250 30S ribosomal prot  
34 36 52.9 1007 2 T42219 alpha-mannosidase  
35 36 52.9 1405 2 T40607 probable dna-dirac  
36 36 52.9 1441 2 T00335 hypothetical prote  
37 36 52.9 1563 1 A58881 breast/ovarian can  
38 36 52.9 3623 2 T09456 intrinsic factor-B  
39 35 51.5 84 2 G86327 protein F18O14.23  
40 35 51.5 92 2 JU0239 nuclear matrix pro  
41 35 51.5 127 2 E95298 conserved hypothet  
42 35 51.5 170 2 AC1219 diol dehydrase (di  
43 35 51.5 170 2 AP1572 diol dehydrase (di  
44 35 51.5 186 2 S05560 hypothetical prote  
45 35 51.5 269 2 T16910 hypothetical prote

#### ALIGNMENTS

##### RESULT 1

E84828 Probable WD-40 repeat protein [imported] - Arabidopsis thaliana

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001

C;Accession: E84828

R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; A

euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.

Nature 402, 761-768, 1999

A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A;Reference number: A84420; MUID:20083487; PMID:10617197

A;Accession: E84828

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-775 <S>

A;Cross-references: GB:AE002093; NID:g4586061; PIDN:AAD25679.1; GSPDB:GN00139

C;Genetics:

A;Gene: At2g40360

A;Map position: 2

Query Match 60.3%; Score 41; DB 2; Length 775;  
Best Local Similarity 66.7%; Pred. No. 21;  
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 DVRPYRTSRYLE 12

Db 429 DURPYNSCYLE 440

##### RESULT 2

G91067 Hypothetical protein ECs3511 [imported] - Escherichia coli (strain O157:H7, substrain RIN

C;Species: Escherichia coli

C;Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 18-Jul-2001

C;Accession: G91067

R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.;

gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.

DNA Res. 8, 11-22, 2001

A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genon

A;Reference number: A98629; MUID:21156231; PMID:11258796

A;Accession: G91067

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-437 <HAY>

A;Cross-references: GB:BA000007; PIDN:BA836934.1; PID:g13362982; GSPDB:GN00154

A;Experimental source: strain O157:H7, substrain RIMD 0509952

C;Genetics:

A;Gene: ECs3511

Query Match 58.8%; Score 40; DB 2; Length 437;  
Best Local Similarity 61.5%; Pred. No. 18;

Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 DVPRPRTSYRLEL 13  
 ||| ||| : |||  
 Db 317 DVRSERTNSYIEL 329

RESULT 3  
 H85911  
 hypothetical protein Z3943 [imported] - Escherichia coli (strain O157:H7, substrain EDL958)  
 C:Species: Escherichia coli  
 C>Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001  
 C:Accession: H85911  
 R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew, M.W.; Lin, A.; Dimalanta, E.; Potamousis, K.; Apodaca, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; J. et al.  
 A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
 A:Reference number: A85480; MUID:21074935; PMID:11206551  
 A:Accession: H85911  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-441 <STO>  
 A:Cross-references: GB:AE005174; NID:gi2517076; PIDN:AAGS7756.1; GSPDB:GN00145; UWGP:Z3943  
 A:Experimental source: strain O157:H7, substrain EDL958  
 C:Genetics:  
 A:Gene: Z3943

Query Match 58.8%; Score 40; DB 2; Length 441;  
 Best Local Similarity 61.5%; Pred. No. 18;  
 Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 DVPRPRTSYRLEL 13  
 ||| ||| : |||  
 Db 321 DVRSERTNSYIEL 333

RESULT 4  
 AD3042  
 oxidoreductase Atu3952 [imported] - Agrobacterium tumefaciens (strain C58, Dupont)  
 C:Species: Agrobacterium tumefaciens  
 C>Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 18-Nov-2002  
 C:Accession: AD3042  
 R:Wood, D.W.; Secubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.B.; Chen, Y.; Woo, I.; Karp, P.; Romero, P.; Zhang, S.; Science 294, 2317-2323, 2001  
 A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, J.E.  
 A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.  
 A:Reference number: AB2577; MUID:21608550; PMID:11743193  
 A:Accession: AD3042  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-442 <KUR>  
 A:Cross-references: GB:AE008689; PIDN:RAL44754.1; PID:gi7742390; GSPDB:GN00187  
 A:Experimental source: strain C58 (Dupont)  
 C:Genetics:  
 A:Gene: Atu3952  
 A:Map position: linear chromosome

Query Match 58.8%; Score 40; DB 2; Length 442;  
 Best Local Similarity 58.3%; Pred. No. 18;  
 Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 DVPRPRTSYRLEL 12  
 ||| ||| : |||  
 Db 421 DPRPYRSRFFD 432

RESULT 5  
 H98243  
 agas protein [imported] - Agrobacterium tumefaciens (strain C58, Cereon)  
 C:Species: Agrobacterium tumefaciens

C>Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 18-Nov-2002  
 C:Accession: H98243  
 R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quorllo, B.; Goldman, A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.; Science 294, 2323-2328, 2001  
 A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tumefaciens  
 A:Reference number: A97359; MUID:21608551; PMID:11743194  
 A:Accession: H98243  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-442 <KUR>  
 A:Cross-references: GB:AE007870; PIDN:AAK89474.1; PID:gi15159342; GSPDB:GN00170  
 C:Genetics:  
 A:Gene: AGR\_L1801  
 A:Map position: linear chromosome

Query Match 58.8%; Score 40; DB 2; Length 442;  
 Best Local Similarity 58.3%; Pred. No. 18;  
 Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 DVPRPRTSYRLEL 12  
 ||| ||| : |||  
 Db 421 DPRPYRSRFFD 432

RESULT 6  
 T00447  
 hypothetical protein T14N5.4 - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C>Date: 01-Feb-1999 #sequence\_revision 01-Feb-1999 #text\_change 22-Oct-1999  
 C:Accession: T00447  
 R:Federspiel, N.A.; Palm, C.J.; Conway, A.B.; Kurtz, D.B.; Conway, A.R.; Au, M.; Araujo, J.; Vyotskaia, V.S.; Yu, G.; Ecker, J.; Theologis, A.; Davis, R.W.  
 A:Submitted to the EMBL Data Library, September 1998  
 A:Reference number: Z14152  
 A:Accession: T00447  
 A:Status: translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-258 <FED>  
 A:Cross-references: EMBL:AC004260; NID:gi3176694; PID:gi3540203; GSPDB:GN00059; ATSP:T14N5.4  
 A:Experimental source: cultivar Columbia  
 C:Genetics:  
 A:Gene: AtSP:T14N5.4  
 A:Map position: 1  
 A:Introns: 38/3; 106/1

Query Match 57.4%; Score 39; DB 2; Length 258;  
 Best Local Similarity 66.7%; Pred. No. 16;  
 Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 4 PYRTSYRLEL 12  
 ||| ||| : |||  
 Db 199 PWRTSRVYMQ 207

RESULT 7  
 T47356  
 hypothetical protein F18P9.140 - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C>Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 20-Apr-2000  
 C:Accession: T47356  
 R:Nyakatura, G.; Fartmann, B.; Dauner, D.; Sterr, W.; Holland, R.; Weichselgartner, M.; Mayer, K.F.X.  
 A:Submitted to the Protein Sequence Database, April 2000  
 A:Reference number: Z24458  
 A:Accession: T47356  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-303 <NVA>  
 A:Cross-references: EMBL:AL138654  
 A:Experimental source: cultivar Columbia; BAC clone F18P9  
 C:Genetics:  
 A:Map position: 3

A; Introns: 47/3; 103/3; 140/2; 182/1; 210/2; 241/3  
 A; Note: F18P9.140

Query Match 57.4%; Score 39; DB 2; Length 303;

Best Local Similarity 75.0%; Pred. No. 19;

Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DVPRYRTS 8

Db 10 DIRPYKTS 17

#### RESULT 8

A86274

C; Species: Arabidopsis thaliana

C; Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 09-Nov-2001

C; Accession: A86274

R; Theologos, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, J.; Chn, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; Ansen, N.F.; Hughes, B.; Huizar, L.

Nature 408, 816-820, 2000

A; Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A; Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, Ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A; Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A; Reference number: A86141; MUID:21016719; PMID:11130712

A; Accession: A86274

A; Status: Preliminary

A; Molecule type: DNA

A; Residues: 1-526 <STO>

A; Cross-references: GB:AB005172; NID:95080782; PIDN:AD39292.1; GSPDB:GN00141

C; Genetics:

A; Map position: 1

Query Match

Best Local Similarity 57.4%; Score 39; DB 2; Length 526;

Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 4 PYRTSRYL 11

Db 79 PYRTSEYL 86

#### RESULT 9

B82559

C; Species: Xylella fastidiosa

C; Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Sep-2000

C; Accession: B82559

R; Anonymous, the Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen

Nature 406, 151-157, 2000

A; Title: The genome sequence of the plant pathogen Xylella fastidiosa.

A; Reference number: A82515; MUID:20365717; PMID:10910347

A; Note: for a complete list of authors see reference number A59328 below

A; Accession: B82559

A; Status: Preliminary

A; Molecule type: DNA

A; Residues: 1-569 <STM>

A; Cross-references: GB:AB004052; GB:AB003849; NID:99107617; PIDN:AAF85237.1; GSPDB:GN001

A; Experimental source: strain 9a5c

R; Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A

Brienes, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carter, H

as-Neto, E.; Docena, C.; El-Dorzy, H.; Facincani, A.P.; Ferreira, A.J.S.

submitted to GenBank, June 2000

A; Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm

J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laig

Chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E

A; Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.

F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A

Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak

A; Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveira  
 M.; Tuhako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z

A; Reference number: A59328

A; Contents: annotation

C; Genetics:

A; Gene: Xf2438

C; Superfamily: Escherichia coli ribosomal protein S1

Query Match 57.4%; Score 39; DB 2; Length 569;

Best Local Similarity 66.7%; Pred. No. 36;

Matches 8; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 DVPRYRTSRYL 12

Db 149 DVPRVDRPAYLE 160

#### RESULT 10

SNCPPI

pitrilysin (BC 3.4.24.55) precursor [validated] - Escherichia coli (strain K-12)

N; Alternate names: endopeptidase Pi; proteinase III

C; Species: Escherichia coli

C; Date: 31-Mar-1993 #sequence\_revision 31-Oct-1997 #text\_change 27-Oct-2003

C; Accession: F65064; A29093; A25765; B25532

R; Blatner, P.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Col

A.; Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A; Title: The complete genome sequence of Escherichia coli K-12.

A; Reference number: A64720; MUID:97426617; PMID:9278503

A; Accession: F65064

A; Status: nucleic acid sequence not shown; translation not shown

A; Molecule type: DNA

A; Residues: 1-962 <BLAT>

A; Cross-references: GB:AE000365; GB:U00096; NID:92367163; PIDN:AA075860.1; PID:G2367164;

A; Experimental source: strain K-12, substrain MG1655

R; Claverie-Martin, F.; Diaz-Torres, M.R.; Kushner, S.R.

Gene 54, 185-195, 1987

A; Title: Analysis of the regulatory region of the protease III (ptr) gene of Escherichia

A; Reference number: A29093; MUID:88005781; PMID:3308636

A; Accession: A29093

A; Molecule type: DNA

A; Residues: 1-276, 'HYHSRL', 283, 'W', 285-296 <CIA>

A; Cross-references: GB:M17095; NID:9147390; PIDN:AAA24436.1; PID:G147391

A; Experimental source: strain K12

A; Note: part of this sequence, including the amino end of the mature protein, was confirm

R; Finch, P.W.; Wilson, R.E.; Brown, K.; Hickson, I.D.; Emmerson, P.T.

Nucleic Acids Res. 14, 7695-7703, 1986

A; Title: Complete nucleotide sequence of the Escherichia coli ptr gene encoding protease

A; Reference number: A25765; MUID:87040734; PMID:3534791

A; Accession: A25765

A; Molecule type: DNA

A; Residues: 1-962 <PIN>

A; Cross-references: GB:X06227; NID:942560; PIDN:CAA29576.1; PID:G42561

R; Becker, A.B.; Roth, R.A.

Proc. Natl. Acad. Sci. U.S.A. 89, 3835-3839, 1992

A; Title: An unusual active site identified in a family of zinc metalloendopeptidases.

A; Reference number: A38854; MUID:92237263; PMID:1570301

A; Contents: annotation; active site

C; Genetics:

A; Gene: ptr

A; Map position: 61

C; Function:

A; Description: endopeptidase degrades small peptides [validated, MUID:92237263]

A; Pathway: protein degradation

C; Superfamily: insulin-degrading enzyme (IDE)

C; Keywords: hydrolase; metalloproteinase; monomer; periplasmic space; protein degradation

F; 1-23/Domain: signal sequence #status experimental <SIG>

F; 24-962/Product: pitrilysin #status experimental <SIG>

F; 88,92/Binding site: zinc (His) #status experimental

F; 91/Active site: Glu #status experimental

Query Match

Best Local Similarity 57.4%; Score 39; DB 1; Length 962;

Matches 70.0%; Pred. No. 61;

```
Matches      7;  Conservative      2;  Mismatches      1;  Indels      0;  Gaps      0;
```

QY	4	PYTSRYLEL	13
		:    :	
Db	124	PYRTAFYLEV	133

RESULT 11

F91088  
proteinase III [imported] - Escherichia coli (strain O157:H7, substrain RMD 0509952)  
C:Species: Escherichia coli  
C:Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 27-Oct-2003  
C:Accession: F91088  
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.;  
Gasaawa, N.; Yasunaga, T.; Kihara, S.; Shiba, T.; Hattori, M.; Shingawa, H.  
DNA Res. 8, 11-22, 2001  
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gene  
A:Reference number: A93629; MUID:21156231; PMID:11258796  
A:Accession: F91088  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-962 -SHAY>  
A:Cross-references: PIDN:BA837101.1; PID:G13363150; GSPDB:GN00154  
A:Experimental source: strain O157:H7, substrain RMD 0509952  
C:Genetics:  
A:Gene: ECs3678  
C:Superfamily: insulin-degrading enzyme (IDE)

QY 4 PYRTSRYLEL 13  
124 PYRTAFYLEV 133

RESULT 12

H85933  
 Phosphatase III [imported] - Escherichia coli (strain O157:H7, substrain EDL933)  
 C.Species: Escherichia coli  
 C.Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #ext\_change 27-Oct-2003  
 C.Accession: H85933  
 R.Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew  
 Miller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,  
 Nature 409, 529-533, 2001  
 A.Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
 A.Reference number: A85480; MUID:21074935; PMID:11208551  
 A.Accession: H85933  
 A.Status: preliminary  
 A.Molecule type: DNA  
 A.Residues: 1-962 <STO>  
 A.Cross-references: GB:AE0051174; NID:ig12517303; PIDN:AA057932.1; GSPDB:GNC0145; UWGPG:Z41  
 A.Experimental source: strain O157:H7, substrain EDL933  
 C.Genetics:  
 A.Gene: ptr  
 A.Superfamily: insulin-degrading enzyme (IDE)

QY	4	PYR	SR	YLE	13
			:		:
Db	124	PYR	AF	YLE	133

RESULT 13

AC0865 protease III precursor (pitrylsin) [imported] - *Salmonella enterica* ser C:species: *Salmonella enterica* subsp. *enterica* serovar Typhi  
A:Note: this species has also been called *Salmonella typhi*  
C:Idate: 09-Nov-2001 #sequence revision 09-Nov-2001 #text change 27-Oct-2003

C/Accession: AC0865  
R/Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moule, S.; O'Gaora, P.

Qy 4 PYRTSRYLEL 13  
124 PYRTAFYLEV 133

RESULT 14

T24300  
ribosomal protein S14 T01E8.6, mitochondrial [similarity] - Caenorhabditis elegans  
CSpecies: Caenorhabditis elegans  
C.Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 03-May-2002  
C.Accession: T24300  
R:McMurray, A.  
submitted to the EMBL Data Library, March 1995  
A.Reference number: Z19871  
A.Accession: T24300  
A.Status: preliminary; translated from GB/EMBL/DBJ  
A.Molecule type: DNA  
A.Residues: 1-103 <MIL>  
A.Cross-references: EMBL:Z48809; PIDN:CAA88746.1; GSPDB:GN00020; CESP:T01E8.6  
A.External source: clone T01E8

```
Qy      2 VRPYRTSYL 11
        ::||| |||:|
Db      76 IKPYRLSRHL 85
```

RESULT 15

H96974  
cyclic\_beta 1-2 glucan synthetase [imported] - Clostridium acetobutylicum  
CSpecies: Clostridium acetobutylicum  
CDate: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 14-Sep-2001  
CAccession: H96974  
R.Nolling, J.; Bretton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee, D.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.  
J. Bacteriol. 183, 4823-4838, 2001  
A.Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clostridium acetobutylicum  
A.Reference number: A56900; MUID:21359325; PMID:21359325  
A.Accession: H96974  
A.Status: preliminary  
A.Molecule type: DNA  
A.Residues: 1-2870 <KUR>  
A.Cross-references: GB:AB001437; PID:gl5023480; GSPDB:GN00168  
A.Experimental source: Clostridium acetobutylicum ATCC824



```

C;Genetics:
A;Gene: CAC0609
Query Match      55.9%; Score 38; DB 2; Length 2870;
Best Local Similarity 77.8%; Pred. NO. 2.8e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      5 YRTSRYLEL 13
      ||: |||||
Db      2594 YRSGRYLEL 2602

Search completed: October 5, 2004, 08:16:50
Job time : 4.10648 secs

```

This Page Blank (uspto)

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2004, 07:29:35 ; Search time 1.16358 Seconds  
(without alignments)  
581.749 Million cell updates/sec

Title: US-09-805-290A-9

Perfect score: 68

Sequence: 1 DVRPYRTSYLEL 13

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	39	57.4	526	1	FUT7 ARATH
2	39	57.4	962	1	PTRA_ECO57
3	39	57.4	962	1	PTRA_ECOL6
4	39	57.4	962	1	PTRA_ECOLI
5	39	57.4	962	1	PTRA_SALTI
6	39	57.4	962	1	PTRA_SALTY
7	39	57.4	962	1	PTRA_SHIFL
8	38	55.9	99	1	DEFC_AEDAE
9	38	55.9	199	1	RT14_CAEEL
10	38	55.9	653	1	MALO_PVRKO
11	38	55.9	3503	1	DS_BROME
12	37	54.4	620	1	SYR_BIFLO
13	36	52.9	378	1	RS1_PROSP
14	36	52.9	396	1	NETK_PSEAE
15	36	52.9	415	1	YBDG_ECOLI
16	36	52.9	432	1	PTNI_MOUSE
17	36	52.9	432	1	PTNI_RAT
18	36	52.9	1007	1	M2BI_FELCA
19	36	52.9	1405	1	RPCI_SCHPO
20	36	52.9	1863	1	BRCI_HUMAN
21	35	51.5	210	1	Y14F_BPT4
22	35	51.5	230	1	FLPA_THEAC
23	35	51.5	290	1	BET4_YEAST
24	35	51.5	434	1	PTNI_CHICK
25	35	51.5	435	1	PTNI_HUMAN
26	35	51.5	443	1	CBPM_HUMAN
27	35	51.5	444	1	V111_FOWPV
28	35	51.5	500	1	FUT6_ARATH
29	35	51.5	532	1	GSI_NEUCR
30	35	51.5	557	1	RS1_ECOLI
31	35	51.5	566	1	RS1_EUCBP
32	34	50.0	157	1	LOXS_BORBU
33	34	50.0	282	1	KD12_RHIME

34	34	50.0	286	1	DAP2_CLOAB	Q97d80	clostridium
35	34	50.0	286	1	YGBB_ECOLI	P11666	escherichia
36	34	50.0	300	1	DHAA_MYCBO	Q8xb14	mycobacteri
37	34	50.0	300	1	DHAA_MYCTU	Q50642	mycobacteri
38	34	50.0	312	1	YORA_BPPI	Q06262	bacterioph
39	34	50.0	321	1	YBI_CHICK	Q06066	gallus gall
40	34	50.0	461	1	TRME_HELPF	Q92j96	helicobacte
41	34	50.0	461	1	TRME_HELPF	Q95991	helicobacte
42	34	50.0	482	1	UXAB_CLOAB	Q97167	clostridium
43	34	50.0	484	1	PER2_VOLCA	P81132	volvox cart
44	34	50.0	487	1	SYE2_THEWA	Q9x218	thermotoga
45	34	50.0	550	1	SYR_MYCLE	P45840	mycobacteri

#### ALIGNMENTS

#### RESULT 1

ID	FUT7 ARATH	STANDARD;	PRT;	526 AA.
AC	Q9XI81; Q9LMF2;			
DT	28-FEB-2003 (Rel. 41, Created)			
DT	28-FEB-2003 (Rel. 41, Last sequence update)			
DT	28-FEB-2003 (Rel. 41, Last annotation update)			
DE	Probable fucosyltransferase 7 (EC 2.4.1.-) (ALFUT7).			
GN	FUT7 OR ATG14070 OR F7A19.15 OR F16A14.19 OR F16A14.28.			
OS	Arabidopsis thaliana (Mouse-ear cress).			
OC	Eukaryota; Viridiplantae; Streptophyta;			
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;			
OC	eurosid II; Brassicales; Brassicaceae; Arabidopsi.			
OX	NCBI_TaxID=3702;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=cv. Columbia;			
RX	MEDLINE=21016719; PubMed=11130712;			
RA	Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S., White O., Alonso J., Altafi H., Araujo R., Bowman C.L., Brooks S.Y., Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W., Dewart K., Chung P., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K., Dunn P., Egtu P., Feldblyum T.V., Feng J.-D., Fong B., Fujii C.Y., Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L., Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E., Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B., Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P., Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziali A., Militscher J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I., Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D., Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M., Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D., Utterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M., Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.; "Sequence and analysis of chromosome 1 of the plant Arabidopsis thaliana.";			
RT	Nature 408:816-820(2000).			
RL				
RN	[2]			
RP	IDENTIFICATION AS PUTATIVE FUCOSYLTRANSFERASE, AND TISSUE SPECIFICITY.			
RX	MEDLINE=21608393; PubMed=11743104;			
RA	Sarria R., Wagner T.A., O'Neill M.A., Faik A., Wilkerson C.G., Keestra K., Rakhel N.V.;			
RT	"Characterization of a family of Arabidopsis genes related to xyloglucan fucosyltransferase1.";			
RL	Plant Physiol. 127:1595-1606(2001).			
CC	-!- FUNCTION: May be involved in cell wall biosynthesis. May act as a fucosyltransferase.			
CC	-!- PATHWAY: Glycosylation.			
CC	-!- SUBCELLULAR LOCATION: Type II membrane protein. Membrane-bound form in trans cisternae of Golgi (By similarity).			
CC	-!- TISSUE SPECIFICITY: Expressed in roots, leaves, stems and seedlings.			
CC	-!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 37.			
CC	-!- CAUTION: Ref.1 (AAF79408) sequence differs from that shown due to erroneous gene model prediction.			

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; AC007576; RAD39292.1; --  
 DR EMBL; AC068197; AAF79408.1; ALT\_SEQ.  
 DR PIR; A86274; A86274.  
 DR InterPro; IPR004938; XG\_Ftase.  
 DR Pfam; PF03254; XG\_Ftase; 1.  
 KW Transferase; Glycosyltransferase; Transmembrane; Glycoprotein;  
 KW Signal-anchor; Golgi stack; Cell wall.  
 FT DOMAIN 1 4  
 FT TRANSMEM 5 25  
 FT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)  
 FT (POTENTIAL).  
 FT LUMENAL, CATALYTIC (POTENTIAL).  
 FT CARBOHYD 211 211 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 215 215 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 363 363 N-LINKED (GLCNAC. .) (POTENTIAL).  
 SQ SEQUENCE 526 AA; 60537 MW; 8F3B447551025FCD CRC64;

Query Match 57.4%; Score 39; DB 1; Length 526;  
 Best Local Similarity 87.5%; Pred. No. 14;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 4 PYRTSRYL 11  
 |||||  
 Db 79 PYRTSEYL 86

RESULT 2  
 ID PTR\_A EC057 STANDARD; PRT; 962 AA.  
 AC Q8X6M8;  
 DT 10-OCT-2003 (Rel. 42, Created)  
 DT 10-OCT-2003 (Rel. 42, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Protease III precursor (EC 3.4.24.55) (Pitrilysin) (Protease pi).  
 GN PTR\_A OR PTR OR Z4138 OR EC83678.  
 OS Escherichia coli O157:H7.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Escherichia.  
 OX NCBI\_TaxID=83334;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=O157:H7 / EDL933 / ATCC 700927;  
 RX MEDLINE=21074935; PubMed=11206551;  
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,  
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,  
 RA Posral G., Hackett J., Klink S., Boutin A., Shao Y., Miller J.,  
 RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamocousis K.,  
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,  
 RA Welch R.A., Blattner F.R.;  
 RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7";  
 RL Nature 409:529-533(2001).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=O157:H7 / RIMD 0509952;  
 RX MEDLINE=21156331; PubMed=11258796;  
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Murata T., Tanaka M., Tobe T.,  
 RA Han C.-G., Ohtsuko E., Nakayama K., Ogasawara N., Yasunaga T.,  
 RA Iida T., Takami H., Honda T., Sasakawa C.,  
 RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;  
 RT "Complete genome sequence of enterohaemorrhagic Escherichia coli  
 O157:H7 and genomic comparison with a laboratory strain K-12";  
 RL DNA Res. 8:11-22(2001).  
 CC -!- FUNCTION: Endopeptidase that degrades small peptides of less than  
 CC 7 kDa, such as glucagon and insulin (By similarity).  
 CC -!- CATALYTIC ACTIVITY: Preferential cleavage of 16-Tyr-|-Leu-17 and  
 CC 25-Phe-|-Tyr-26 bonds of oxidized insulin B chain. Also acts on

CC other substrates of Mw less than 7 kDa such as insulin and  
 CC glucagon.  
 CC -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).  
 CC -!- SUBUNIT: Monomer (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Periplasmic (By similarity).  
 CC -!- SIMILARITY: Belongs to peptidase family M16.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; AE005510; AAG57932.1; --  
 DR EMBL; AP002563; BAB37101.1; --  
 DR PIR; F91088; F91088.  
 DR InterPro; IPR001431; Peptidase\_M16.  
 DR InterPro; IPR007863; Peptidase\_M16\_C.  
 DR Pfam; PF00675; Peptidase\_M16; 1.  
 DR Pfam; PF05193; Peptidase\_M16\_C; 2.  
 DR PROSITE; PS00143; INSULINASE; 1.  
 KW Hydrolase; Metalloprotease; Magnesium; Periplasmic; Zinc; Signal;  
 KW Complete proteome.  
 FT SIGNAL 1 23 BY SIMILARITY.  
 FT CHAIN 24 962 PROTEASE III.  
 FT METAL 88 88 ZINC (BY SIMILARITY).  
 FT ACT\_SITE 91 91 BY SIMILARITY.  
 FT METAL 92 92 ZINC (BY SIMILARITY).  
 FT METAL 169 169 ZINC (BY SIMILARITY).  
 SQ SEQUENCE 962 AA; 107808 MW; 30785644CBF0E452 CRC64;

Query Match 57.4%; Score 39; DB 1; Length 962;  
 Best Local Similarity 70.0%; Pred. No. 26;  
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 4 PYRTSRYLEL 13  
 |||||  
 Db 124 PYRTAFYLEV 133

RESULT 3  
 ID PTR\_A EC06 STANDARD; PRT; 962 AA.  
 AC Q8CVS2;  
 DT 10-OCT-2003 (Rel. 42, Created)  
 DT 10-OCT-2003 (Rel. 42, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Protease III precursor (EC 3.4.24.55) (Pitrilysin) (Protease pi).  
 GN PTR\_A OR PTR OR C3415.  
 OS Escherichia coli O6.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Escherichia.  
 OX NCBI\_TaxID=217992;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=O6:H1 / CFT073 / ATCC 700928;  
 RX MEDLINE=22388234; PubMed=12471157;  
 RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,  
 RA Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,  
 RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,  
 RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;  
 RT "Extensive mosaic structure revealed by the complete genome sequence  
 of uropathogenic Escherichia coli";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).  
 CC -!- FUNCTION: Endopeptidase that degrades small peptides of less than  
 CC 7 kDa, such as glucagon and insulin (By similarity).  
 CC -!- CATALYTIC ACTIVITY: Preferential cleavage of 16-Tyr-|-Leu-17 and  
 CC 25-Phe-|-Tyr-26 bonds of oxidized insulin B chain. Also acts on  
 CC other substrates of Mw less than 7 kDa such as insulin and  
 CC glucagon.  
 CC -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).

CC -1- SUBUNIT: Monomer (By similarity).  
 CC -1- SUBCELLULAR LOCATION: Periplasmic (By similarity).  
 CC -1- SIMILARITY: Belongs to peptidase family M16.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; AB016765; AAN81860.1; -  
 CC InterPro; IPR001431; Peptidase\_M16.  
 CC InterPro; IPR007863; Peptidase\_M16\_C.  
 CC Pfam; PF00675; Peptidase\_M16; 1.  
 CC Pfam; PF05193; Peptidase\_M16\_C; 2.  
 CC PROSITE; PS00143; INSULINASE; 1.  
 CC Hydrolyase; Metalloprotease; Magnesium; Zinc; Signal;  
 KW Complete proteome.  
 FT SIGNAL 1 23 BY SIMILARITY.  
 FT CHAIN 24 962 PROTEASE III.  
 FT METAL 88 88 ZINC (BY SIMILARITY).  
 FT ACT\_SITE 91 91 BY SIMILARITY.  
 FT METAL 92 92 ZINC (BY SIMILARITY).  
 FT METAL 169 169 ZINC (BY SIMILARITY).  
 SQ SEQUENCE 962 AA; 107891 MW; DEDD2CA2A9ADF8D CRC64;  
 Query Match 57.4%; Score 39; DB 1; Length 962;  
 Best Local Similarity 70.0%; Pred. No. 26;  
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 QY 4 PYRTSRYLEL 13  
 DB 124 PYRTAFYLEV 133  
 RESULT 4  
 PTRA ECOLI  
 ID -PTRA ECOLI STANDARD; PRT; 962 AA.  
 AC P05458; P78106;  
 DT 01-NOV-1988 (Rel. 09, Created)  
 DT 01-NOV-1988 (Rel. 09, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Protease III precursor (EC 3.4.24.55) (Pitriylsin) (Protease pi).  
 GN PTRA OR PTR OR B2821.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Escherichia.  
 OX NCBI\_Taxid=562;  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RX MEDLINE=87040734; PubMed=3534791;  
 RA Finch P.W., Wilson R.E., Brown K., Hickson I.D., Emerson P.T.;  
 FT "Complete nucleotide sequence of the Escherichia coli ptr gene  
 encoding protease III";  
 RL Nucleic Acids Res. 14:7695-7703(1986).  
 RN [2]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=K12 / MG1655;  
 RX MEDLINE=97426617; PubMed=9278503;  
 RA Blattner F.R., Plunkett G. III, Aloch C.A., Perna N.T., Burland V.,  
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.P.,  
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 RA Mau B., Shao Y.;  
 RT "The complete genome sequence of Escherichia coli K-12";  
 RL Science 277:1453-1474(1997).  
 RN [3]  
 RN SEQUENCE OF 853-962 FROM N.A.  
 RX MEDLINE=87068729; PubMed=3537960;  
 RA Finch P.W., Storey A., Chapman K.E., Brown K., Hickson I.D.,  
 RA Emerson P.T.;  
 RT "Complete nucleotide sequence of the Escherichia coli recB gene";

RL Nucleic Acids Res. 14:8573-8582(1986).  
 RN [4]  
 RN SEQUENCE OF 1-296 FROM N.A.  
 RC STRAIN=K12;  
 RX MEDLINE=88005781; PubMed=3308636;  
 RA Claverie-Martin F., Diaz-Torres M.R., Kushner S.R.;  
 RT "Analysis of the regulatory region of the protease III (ptr) gene of  
 Escherichia coli K-12";  
 RL Gene 54:185-195(1987).  
 RN [5]  
 RN MUTAGENESIS, AND ACTIVE SITE.  
 RX MEDLINE=92237263; PubMed=1570301;  
 RA Becker A.B., Roth R.A.;  
 RT "An unusual active site identified in a family of zinc  
 metalloendopeptidases";  
 RL Proc. Natl. Acad. Sci. U.S.A. 89:3835-3839(1992).  
 RN [6]  
 RN MUTAGENESIS, AND ACTIVE SITE.  
 RX MEDLINE=93277449; PubMed=8099278;  
 RA Becker A.B., Roth R.A.;  
 RT "Identification of glutamate-169 as the third zinc-binding residue in  
 proteinase III, a member of the family of insulin-degrading  
 enzymes";  
 RL Biochem. J. 292:137-142(1993).  
 CC -1- FUNCTION: Endopeptidase that degrades small peptides of less than  
 7 kDa, such as glucagon and insulin.  
 CC -1- CATALYTIC ACTIVITY: Preferential cleavage of 16-Tyr-Leu-17 and  
 25-Phe-Tyr-26 bonds of oxidized insulin B chain. Also acts on  
 other substrates of Mw less than 7 kDa such as insulin and  
 glucagon.  
 CC -1- COFACTOR: Binds 1 zinc ion per subunit.  
 CC -1- SUBUNIT: Monomer.  
 CC -1- SUBCELLULAR LOCATION: Periplasmic.  
 CC -1- SIMILARITY: Belongs to peptidase family M16.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; X04581; CAA28249.1; -  
 CC EMBL; U29581; AAB40468.1; -  
 CC EMBL; AF000365; AAC75860.1; -  
 CC EMBL; X06227; CAA29576.1; -  
 CC EMBL; M17095; AAA24436.1; -  
 CC PIR; F65064; SNECFI.  
 CC MEROPS; M16.001; 6TH EDITION.  
 CC ECODBASE; G095.0; 6TH EDITION.  
 CC Ecogene; EG10786; ptrA.  
 CC InterPro; IPR001431; Peptidase\_M16.  
 CC InterPro; IPR007863; Peptidase\_M16\_C.  
 CC Pfam; PF00675; Peptidase\_M16; 1.  
 CC Pfam; PF05193; Peptidase\_M16\_C; 2.  
 CC PROSITE; PS00143; INSULINASE; 1.  
 CC Hydrolyase; Metalloprotease; Magnesium; Zinc; Signal;  
 KW Complete proteome.  
 FT SIGNAL 1 23  
 FT CHAIN 24 962 PROTEASE III.  
 FT METAL 88 88 ZINC.  
 FT ACT\_SITE 91 91  
 FT METAL 92 92 ZINC.  
 FT METAL 169 169 ZINC.  
 FT MUTAGEN 88 88 H->R: LOSS OF ACTIVITY AND OF ZN-BINDING.  
 FT MUTAGEN 91 91 E->Q: LOSS OF ACTIVITY.  
 FT MUTAGEN 92 92 H->R: LOSS OF ACTIVITY AND OF ZN-BINDING.  
 FT MUTAGEN 162 162 E->Q: 20% LOSS OF ACTIVITY.  
 FT MUTAGEN 169 169 E->Q: LOSS OF ACTIVITY AND OF ZN-BINDING.  
 FT MUTAGEN 204 204 E->Q: NO LOSS OF ACTIVITY.  
 FT CONFLICT 277 284 ITHKVPK -> HYSLRPW (IN REF. 4).  
 SQ SEQUENCE 962 AA; 107708 MW; 0558C68C2F1A0540 CRC64;

Query Match 57.4%; Score 39; DB 1; Length 962;  
 Best Local Similarity 70.0%; Pred. No. 26;  
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 4 PYRTSRYLEL 13  
 DB 124 PYRTAFYLEV 133

## RESULT 5

PTSA\_SALTY STANDARD; PRT; 962 AA.  
 ID PTSA\_SALTY STANDARD; PRT; 962 AA.  
 AC Q8Z418;  
 DT 10-OCT-2003 (Rel. 42, Created)  
 DT 10-OCT-2003 (Rel. 42, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Protease III precursor (EC 3.4.24.55) (Pitriylsin) (Protease pi).  
 GN PTSA OR PTR OR STY3133 OR T2903.  
 OS Salmonella typhi.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Salmonella.  
 OX NCBI\_TaxID=601;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CT18;  
 RX MEDLINE=21534947; PubMed=11677608;  
 RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,  
 RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebaihia M.,  
 RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,  
 RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,  
 RA Fellwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,  
 RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,  
 RA Quail M.A., Rutherford K., Simmonds M., Skelton J., Stevens K.,  
 RA Whitehead S., Barrrell B.G.;  
 RT "Complete genome sequence of a multiple drug resistant Salmonella  
 RT enterica serovar Typhi CT18.";  
 RL Nature 413:848-852(2001).

[2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Ty2 / ATCC 700931;  
 RX MEDLINE=22531367; PubMed=12644504;  
 RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,  
 RA Burland V., Kodoyianni V., Schwartz D.C., Blattner F.R.;  
 RT "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2  
 RT and CT18.";  
 RL J. Bacteriol. 185:2330-2337(2003).  
 CC -!- FUNCTION: Endopeptidase that degrades small peptides of less than  
 CC 7 kDa, such as glucagon and insulin (By similarity).  
 CC -!- CATALYTIC ACTIVITY: Preferential cleavage of 16-Tyr-Leu-17 and  
 CC 25-Phe-Tyr-26 bonds of oxidized insulin B chain. Also acts on  
 CC other substrates of Mw less than 7 kDa such as insulin and  
 CC glucagon.  
 CC -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).  
 CC -!- SUBUNIT: Monomer (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Periplasmic (By similarity).  
 CC -!- SIMILARITY: Belongs to peptidase family M16.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; AL627277; CAD02819.1; -;  
 CC EMBL; AE016843; AA070457.1; -;  
 CC NEROPS; M16.001; -;  
 CC InterPro; IPR001431; Peptidase\_M16.  
 CC Pfam; PF007863; Peptidase\_M16.C.  
 CC Pfam; PF00675; Peptidase\_M16; 1.  
 CC Pfam; PF05193; Peptidase\_M16.C; 2.

DR PROSITE; PS00143; INSULINASE; FALSE NEG.  
 KW Hydrolase; Metalloprotease; Magnesium; Periplasmic; Zinc; Signal;  
 KW Complete proteome.  
 FT SIGNAL 1 23 BY SIMILARITY.  
 FT CHAIN 24 962 PROTEASE III.  
 FT METAL 88 88 ZINC (BY SIMILARITY).  
 FT ACT\_SITE 91 91 BY SIMILARITY.  
 FT METAL 92 92 ZINC (BY SIMILARITY).  
 FT METAL 169 169 ZINC (BY SIMILARITY).  
 SQ SEQUENCE 962 AA; 107524 MW; 5745C2C77F7F5832 CRC64;

Query Match 57.4%; Score 39; DB 1; Length 962;  
 Best Local Similarity 70.0%; Pred. No. 26;  
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 4 PYRTSRYLEL 13  
 DB 124 PYRTAFYLEV 133

## RESULT 6

PTSA\_SALTY STANDARD; PRT; 962 AA.  
 ID PTSA\_SALTY STANDARD; PRT; 962 AA.  
 AC Q8ZM65;  
 DT 10-OCT-2003 (Rel. 42, Created)  
 DT 10-OCT-2003 (Rel. 42, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Protease III precursor (EC 3.4.24.55) (Pitriylsin) (Protease pi).  
 GN PTSA OR PTR OR STM2995.  
 OS Salmonella typhimurium.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Salmonella.  
 OX NCBI\_TaxID=602;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=LT2 / SGSC1412 / ATCC 700720;  
 RX MEDLINE=21534948; PubMed=11677609;  
 RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,  
 RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,  
 RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,  
 RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,  
 RA Waterston R., Wilson R.K.;  
 RT "Complete genome sequence of Salmonella enterica serovar Typhimurium  
 RT LT2.";  
 RL Nature 413:852-856(2001).  
 CC -!- FUNCTION: Endopeptidase that degrades small peptides of less than  
 CC 7 kDa, such as glucagon and insulin (By similarity).  
 CC -!- CATALYTIC ACTIVITY: Preferential cleavage of 16-Tyr-Leu-17 and  
 CC 25-Phe-Tyr-26 bonds of oxidized insulin B chain. Also acts on  
 CC other substrates of Mw less than 7 kDa such as insulin and  
 CC glucagon.  
 CC -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).  
 CC -!- SUBUNIT: Monomer (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Periplasmic (By similarity).  
 CC -!- SIMILARITY: Belongs to peptidase family M16.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; AE008837; AAL21871.1; -;  
 CC StyGene; SG22222; pta.  
 CC InterPro; IPR001431; Peptidase\_M16.  
 CC InterPro; IPR007863; Peptidase\_M16.C.  
 CC Pfam; PF00675; Peptidase\_M16; 1.  
 CC Pfam; PF05193; Peptidase\_M16.C; 2.  
 CC PROSITE; PS00143; INSULINASE; 1.  
 KW Hydrolase; Metalloprotease; Magnesium; Periplasmic; Zinc; Signal;  
 KW Complete proteome.

FT SIGNAL 1 23 BY SIMILARITY.  
 FT CHAIN 24 962 PROTEASE III.  
 FT METAL 88 88 ZINC (BY SIMILARITY).  
 FT ACT SITE 91 91 BY SIMILARITY.  
 FT METAL 92 92 ZINC (BY SIMILARITY).  
 FT METAL 169 169 ZINC (BY SIMILARITY).  
 SQ SEQUENCE 962 AA; 107486 MW; 322AD6587B873952 CRC64;

Query Match 57.4%; Score 39; DB 1; Length 962;  
 Best Local Similarity 70.0%; Pred. No. 26;  
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 4 PYRTSRYLEL 13  
 DB 124 PYRTAFYLEV 133

RESULT 7  
 PTR\_A SHIFL STANDARD; PRT; 962 AA.  
 ID Q83QC3; 2003 (Rel. 42, Created)  
 DT 10-OCT-2003 (Rel. 42, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Protease III precursor (EC 3.4.24.55) (Pitrilysin) (Protease pi).  
 GN PTRA OR PTR OR SF2832 OR S3029.  
 OS Shigella flexneri.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Shigella.  
 OX NCBI\_TaxID=623;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=301 / Serotype 2a;  
 RX MEDLINE=2272406; PubMed=12384590;  
 RA Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,  
 RA Yang J., Yang F., Zhang J., Yang G., Wu H., Qu D., Dong J.,  
 RA Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,  
 RA Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,  
 RA Yu J.;  
 RA "Genome sequence of Shigella flexneri 2a: insights into pathogenicity  
 RT through comparison with genomes of Escherichia coli K12 and O157";  
 RL Nucleic Acids Res. 30:4432-4441(2002).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=2457T / ATCC 700930 / Serotype 2a;  
 RX MEDLINE=22590274; PubMed=12704152;  
 RA Wei J., Goldberg M.B., Burkland V., Venkatesan M.M., Deng W.,  
 RA Fournier G., Mayhew G.F., Plunkett G. III, Rose D.J., Darling A.,  
 RA Mau B., Perna N.T., Payne S.M., Runyen-Janecky L.J., Zhou S.,  
 RA Schwartz D.C., Blattner F.R.;  
 RA "Complete genome sequence and comparative genomics of Shigella  
 RT flexneri serotype 2a strain 2457T";  
 RL Infect. Immun. 71:2775-2786(2003).  
 CC -!- FUNCTION: Endopeptidase that degrades small peptides of less than  
 CC 7 kDa, such as glucagon and insulin (By similarity).  
 CC -!- CATALYTIC ACTIVITY: Preferential cleavage of 16-tyr-|-Leu-17 and  
 CC 25-Phe-|-Tyr-26 bonds of oxidized insulin B chain. Also acts on  
 CC other substrates of Mw less than 7 kDa such as insulin and  
 CC glucagon.  
 CC -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).  
 CC -!- SUBUNIT: Monomer (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Periplasmic (By similarity).  
 CC -!- SIMILARITY: Belongs to peptidase family M16.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; AEO15297; AAN4319.1; -.

DR EMBL; AEO16987; AAP18144.1; -.  
 DR InterPro; IPR001431; Peptidase M16.  
 DR InterPro; IPR007863; Peptidase M16\_C.  
 DR Pfam; PF00675; Peptidase\_M16; 1.  
 DR Pfam; PF05193; Peptidase\_M16\_C; 2.  
 DR PROSITE; PS00143; INSULINASE; 1.  
 KW Hydrolase; Metalloprotease; Magnesium; Zinc; Signal;  
 KW Complete proteome.  
 FT SIGNAL 1 23 BY SIMILARITY.  
 FT CHAIN 24 962 PROTEASE III.  
 FT METAL 88 88 ZINC (BY SIMILARITY).  
 FT ACT SITE 91 91 BY SIMILARITY.  
 FT METAL 92 92 ZINC (BY SIMILARITY).  
 FT METAL 169 169 ZINC (BY SIMILARITY).  
 SQ SEQUENCE 962 AA; 107770 MW; 662041A5DCA0F254 CRC64;

Query Match 57.4%; Score 39; DB 1; Length 962;  
 Best Local Similarity 70.0%; Pred. No. 26;  
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 4 PYRTSRYLEL 13  
 DB 124 PYRTAFYLEV 133

RESULT 8  
 DEFC AEDAE STANDARD; PRT; 99 AA.  
 ID P81603; Q9Y0F0; Q9Y0F1;  
 AC 30-MAY-2000 (Rel. 39, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Defensin C precursor.  
 OS Aedes aegypti (Yellowfever mosquito).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Aedes.  
 OX NCBI\_TaxID=7159;  
 RN [1]  
 RP SEQUENCE FROM N.A.; DEVELOPMENTAL STAGE, INDUCTION, AND SUBCELLULAR  
 RP LOCATION.  
 RC STRAIN=Liverpool; TISSUE=Fat body;  
 RX MEDLINE=99124369; PubMed=9927179;  
 RA Lowenberger C.A., Smartt C.T., Bulet P., Ferdig M.T., Severson D.W.,  
 RA Hoffmann J.A., Christensen B.M.;  
 RA "Insect immunity: molecular cloning, expression, and characterization  
 RT of cDNAs and genomic DNA encoding three isoforms of insect defensin in  
 RL Aedes aegypti";  
 RN Insect Mol. Biol. 8:107-118(1999).  
 RN [2]  
 RP SEQUENCE OF 60-99, AND INDUCTION.  
 RC STRAIN=Liverpool; PubMed=7633471;  
 RX MEDLINE=95360030; PubMed=7633471;  
 RA Lowenberger C., Bulet P., Charlet M., Hetru C., Hodgeman B.,  
 RA Christensen B.M., Hoffmann J.A.;  
 RA "Insect immunity: isolation of three novel inducible antibacterial  
 RT defensins from the vector mosquito, Aedes aegypti";  
 RL Insect Biochem. Mol. Biol. 25:867-873(1995).  
 RN [3]  
 RP SEQUENCE OF 60-99.  
 RC STRAIN=REFM;  
 RX MEDLINE=96047965; PubMed=7568275;  
 RA Chalk R., Albuquerque C.M., Ham P.J., Townson H.;  
 RA "Full sequence and characterization of two insect defensins: immune  
 RT peptides from the mosquito Aedes aegypti";  
 RL Proc. R. Soc. Lond., B, Biol. Sci. 261:217-221(1995).  
 CC -!- FUNCTION: Antibacterial peptide mostly active against Gram-  
 CC positive bacteria.  
 CC -!- SUBCELLULAR LOCATION: Secreted.  
 CC -!- TISSUE SPECIFICITY: Hemolymph.  
 CC -!- DEVELOPMENTAL STAGE: Expressed 30 minutes after infection and  
 CC remained present through to 21 days. Expressed in white or callow  
 CC pupae during metamorphosis, but no expression was seen in larvae.  
 CC -!- INDUCTION: By bacterial infection.

CC -!- POLYMORPHISM: There are two defensin C isoforms, C1 (shown here)  
 CC and C2.  
 CC -!- SIMILARITY: Belongs to the arthropod defensin family. Subfamily 1.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; AF156092; AAD40116.2; -  
 CC EMBL; AF156093; AAD40117.2; -  
 CC HSSP; P10891; 11CA.  
 CC InterPro; IPR001542; Defensin\_anpod.  
 CC InterPro; IPR003614; Knot1.  
 CC Pfam; PF01097; Arthro defensin; 1.  
 CC PRINTS; PR00271; DEFENSIN.  
 CC SMART; SM00505; Knot1.1.  
 CC PROSITE; PS00425; ARTHROPOD\_DEFENSINS; 1.  
 CC Insect immunity; Antibiotic; Defensin; Signal; Polymorphism.  
 CC FT SIGNAL 1 23 POTENTIAL.  
 CC FT PROPEP 24 59  
 CC FT CHAIN 60 99 DEFENSIN C.  
 CC FT DISULFID 62 89 BY SIMILARITY.  
 CC FT DISULFID 75 95 BY SIMILARITY.  
 CC FT DISULFID 79 97 BY SIMILARITY.  
 CC FT VARIANT 26 26 G -> E (IN ISOFORM C2).  
 CC FT VARIANT 34 34 P -> S (IN ISOFORM C2).  
 CC FT CONFLICT 83 83 P -> S (IN REF. 2 AND 3).  
 CC FT CONFLICT 91 91 A -> S (IN REF. 2 AND 3).  
 CC SEQUENCE 99 AA; 10696 MW; 7C3566A68F4548ED CRC64;  
 CC  
 CC Query Match 55.9%; Score 38; DB 1; Length 99;  
 CC Best Local Similarity 61.5%; Pred. No. 3.3;  
 CC Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;  
 CC  
 CC QY 1 DVPRTSRYLEL 13  
 CC ||||| : : : : :  
 CC DB 31 DVPYANSLFDEL 43  
 CC  
 CC RESULT 9  
 CC RT14 CAEEL STANDARD; PRT; 199 AA.  
 CC AC P49391.  
 CC DT 01-FEB-1996 (Rel. 33, Created)  
 CC DT 10-OCT-2003 (Rel. 42, Last sequence update)  
 CC DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 CC DE Putative mitochondrial 40S ribosomal protein S14.  
 CC GN T01B8.6.  
 CC OS Caenorhabditis elegans.  
 CC OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 CC Rhabditidae; Feloderinae; Caenorhabditis.  
 CC OX NCBI\_TaxID=6239;  
 CC RN [1]  
 CC RP SEQUENCE FROM N.A.  
 CC RC STRAIN=Bristol N2;  
 CC RA McMurray A.A.;  
 CC RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.  
 CC RN [2]  
 CC RP REVISIONS.  
 CC RA Durbin R.;  
 CC RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.  
 CC -!- SUBCELLULAR LOCATION: Mitochondrial (By similarity).  
 CC -!- SIMILARITY: Belongs to the S14P family of ribosomal proteins.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; Z49809; CAA88746.2; -  
 CC FIR; I24300; I24300.  
 CC WormPep; T01B8.6; CE32923.  
 CC InterPro; IPR001209; Ribosomal\_S14.  
 CC Pfam; PF00253; Ribosomal\_S14; 1.  
 CC PROSITE; PS00527; RIBOSOMAL\_S14; FALSE\_NEG.  
 CC Ribosomal protein; Mitochondrion.  
 CC SEQUENCE 199 AA; 23414 MW; 2CC0021A618BCFF0 CRC64;  
 CC  
 CC Query Match 55.9%; Score 38; DB 1; Length 199;  
 CC Best Local Similarity 60.0%; Pred. No. 7.2;  
 CC Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
 CC  
 CC QY 2 VRPVRTSRYL 11  
 CC : : : : :  
 CC DB 172 IKPYRLSRHL 181  
 CC  
 CC RESULT 10  
 CC MALQ PYRKO STANDARD; PRT; 653 AA.  
 CC AC Q22450.  
 CC DT 16-OCT-2001 (Rel. 40, Created)  
 CC DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 CC DE 4-alpha-glucanotransferase (EC 2.4.1.25) (Amylomaltase)  
 CC DB (Disproportionating enzyme) (D-enzyme).  
 CC OS Pyrococcus kodakarensis  
 CC OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;  
 CC Thermococcus.  
 CC OX NCBI\_TaxID=69014;  
 CC RN [1]  
 CC RP SEQUENCE FROM N.A.  
 CC RC STRAIN=KOD1;  
 CC RA Tachibana Y, Fujiwara S, Takagi M, Imanaka T.;  
 CC RT "Cloning and expression of the 4-alpha-glucanotransferase gene from  
 CC RT the hyperthermophilic archaeon Pyrococcus sp. KOD1, and  
 CC RT characterization of the enzyme."  
 CC RL J. Ferment. Bioeng. 83:540-548(1997).  
 CC -!- CATALYTIC ACTIVITY: Transfers a segment of a (1,4)-alpha-D-glucan  
 CC to a new 4-position in an acceptor, which may be glucose or (1,4)-  
 CC alpha-D-glucan.  
 CC -!- SIMILARITY: Belongs to family 57 of glycosyl hydrolases.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; D87907; BAA22062.1; -  
 CC InterPro; IPR004300; Glyco\_hydro\_57.  
 CC Pfam; PF03065; Glyco\_hydro\_57; 1.  
 CC KW Transferase; Glycosyltransferase; Carbohydrate metabolism.  
 CC SEQUENCE 653 AA; 76664 MW; B0C4695613F29219 CRC64;  
 CC  
 CC Query Match 55.9%; Score 38; DB 1; Length 653;  
 CC Best Local Similarity 77.8%; Pred. No. 27;  
 CC Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 CC  
 CC QY 5 YRTSRYLEL 13  
 CC ||||| : : : : :  
 CC DB 501 YRLSRYIEL 509  
 CC  
 CC RESULT 11  
 CC DS\_DROME STANDARD; PRT; 3503 AA.  
 CC ID\_DS\_DROME



AC Q24292; O9VFS4;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Dachsous protein precursor (Adherin).  
GN DS OR CGI7941.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A., FUNCTION, TISSUE SPECIFICITY, AND DEVELOPMENTAL  
RP STAGE.  
RP TISSUE=Embryo;  
RC MEDLINE=95324813; PubMed=7601355;  
RX Clark H.F., Brentnup D., Schneitz K., Bieber A., Goodman C., Noll M.,  
RT "Dachsous encodes a member of the cadherin superfamily that controls  
RT imaginal disc morphogenesis in Drosophila."  
RL Genes Dev. 9:1530-1542 (1995).  
RN [2]  
RP REVISIONS.  
RP Noll M.;  
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RP STRAIN=Berkley;  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Pelnik S.E., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA Amanatides P.G., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA George R.A., Wotman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA April J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
RA Butts K.C., Busam D.A., Butler J., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Deicher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Hazis N.I., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclet J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of Drosophila melanogaster."  
RL Science 287:2185-2195 (2000).  
RN [4]  
RP REVISIONS.  
RX MEDLINE=22426069; PubMed=12537572;  
RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,  
RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,  
RA Smith C.D., Tupy J.L., Whitfield B.J., Bayraktaroglu L., Berman B.P.,  
RA Bettencourt B.R., Celisner S.E., de Grey A.D.N.J., Drysdale R.A.,

RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,  
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,  
RA Lewis S.E.;  
RT "Annotation of the Drosophila melanogaster euchromatic genome: a  
RT systematic review."  
RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22 (2002).  
CC -!- FUNCTION: Involved in morphogenesis. May also be involved in cell  
CC adhesion.  
CC -!- SUBCELLULAR LOCATION: Type I membrane protein (Potential).  
CC -!- TISSUE SPECIFICITY: Expressed in embryonic ectoderm. In larvae,  
CC expression is restricted to imaginal disks and brain.  
CC -!- DEVELOPMENTAL STAGE: Expressed throughout embryogenesis where it  
CC is first detected during gastrulation. Also expressed in larvae  
CC and adults.  
CC -!- SIMILARITY: Contains 27 cadherin domains.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: L08811; AAA79329.2; -;  
DR EMBL: AB003588; AAF51468.3; ALT\_INIT.  
DR HSSP: P15116; INCI.  
DR FlyBase: FBgn0000497; ds.  
DR GO: GO:0005887; C:integral to plasma membrane; ISS.  
DR GO: GO:0008014; F:calcium-dependent cell adhesion molecule ac. .; ISS.  
DR GO: GO:0016339; P:calcium-dependent cell-cell adhesion; ISS.  
DR GO: GO:0008283; P:cell proliferation; IMP.  
DR GO: GO:000904; P:cellular morphogenesis during differentiation; IMP.  
DR GO: GO:0045317; P:equator specification; IMP.  
DR GO: GO:0045198; P:reestablishment of epithelial cell polarity; IMP.  
DR GO: GO:0018149; P:protein-protein cross-linking; IPI.  
DR InterPro: IPR000233; Cadherin\_C\_term.  
DR Pfam: PF00028; cadherin; 26.  
DR Pfam: PF01049; Cadherin\_C term; 1.  
DR PRINTS: PR00205; CADHERIN.  
DR SMART: SM00112; CA; 25.  
DR PROSITE: PS00232; CADHERIN\_1; 20.  
DR PROSITE: PS0268; CADHERIN\_2; 27.  
DR Cell adhesion; Glycoprotein; Transmembrane; Calcium; Calcium-binding;  
KW Repeat; Signal; Developmental protein.  
FT SIGNAL 1 20  
FT CHAIN 21 3503 DACHSOUS PROTEIN.  
FT DOMAIN 21 3045 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 3046 3066 POTENTIAL.  
FT DOMAIN 3067 3503 CYTOPLASMIC (POTENTIAL).  
FT DOMAIN 22 121 CADHERIN 1.  
FT DOMAIN 122 233 CADHERIN 2.  
FT DOMAIN 234 340 CADHERIN 3.  
FT DOMAIN 345 451 CADHERIN 4.  
FT DOMAIN 452 558 CADHERIN 5.  
FT DOMAIN 559 662 CADHERIN 6.  
FT DOMAIN 663 774 CADHERIN 7.  
FT DOMAIN 775 878 CADHERIN 8.  
FT DOMAIN 879 983 CADHERIN 9.  
FT DOMAIN 984 1100 CADHERIN 10.  
FT DOMAIN 1101 1203 CADHERIN 11.  
FT DOMAIN 1205 1312 CADHERIN 12.  
FT DOMAIN 1313 1432 CADHERIN 13.  
FT DOMAIN 1433 1549 CADHERIN 14.  
FT DOMAIN 1556 1666 CADHERIN 15.  
FT DOMAIN 1667 1794 CADHERIN 16.  
FT DOMAIN 1796 1899 CADHERIN 17.  
FT DOMAIN 1900 2004 CADHERIN 18.  
FT DOMAIN 2005 2111 CADHERIN 19.  
FT DOMAIN 2114 2269 CADHERIN 20.  
FT DOMAIN 2270 2375 CADHERIN 21.  
FT DOMAIN 2375 2479 CADHERIN 22.

```
FT DOMAIN 2489 2595 CADHERIN 23.
FT DOMAIN 2596 2695 CADHERIN 24.
FT DOMAIN 2701 2809 CADHERIN 25.
FT DOMAIN 2810 2916 CADHERIN 26.
FT DOMAIN 2919 3028 CADHERIN 27.
FT CONFLICT 1070 1070 V -> I (IN REF. 1).
FT CONFLICT 1490 1490 R -> S (IN REF. 1).
FT CONFLICT 1636 1636 G -> S (IN REF. 1).
FT CONFLICT 1692 1692 S -> P (IN REF. 1).
FT CONFLICT 1804 1804 V -> I (IN REF. 1).
FT CONFLICT 2029 2029 L -> I (IN REF. 1).
FT CONFLICT 2210 2210 P -> A (IN REF. 1).
FT CONFLICT 2289 2289 A -> S (IN REF. 1).
FT CONFLICT 2536 2536 S -> T (IN REF. 1).
FT CONFLICT 2862 2862 R -> Q (IN REF. 1).
FT CONFLICT 3038 3038 S -> G (IN REF. 1).
SQ SEQUENCE 3503 AA; 379774 MW; 975B09F059F7EEF5 CRC64;

Query Match 55.9%; Score 38; DB 1; Length 3503;
Best Local Similarity 60.0%; Pred. No. 1.7e+02;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 DVPYPTSYRY 10
   | : | | | |
DB 149 DUEPYNTQRY 158

RESULT 12
SYR_BIFLO
ID SYR_BIFLO STANDARD; PRT; 620 AA.
AC Q8G4V2;
DT 15-MAR-2004 (Rel. 43, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DE Arginyl-tRNA synthetase (EC 6.1.1.19) (Arginine-tRNA ligase) (ArgRS).
GN ARG5 OR BL1272.
OS Bifidobacterium longum.
OC Bacteria; Actinobacteria; Actinobacteridae; Bifidobacteriales;
OC Bifidobacteriaceae; Bifidobacterium.
OX NCBI_TaxID=216816;
RN [1]
RC STRAIN=NCC 2705.
RX MEDLINE=22294977; PubMed=12381787;
RA Schell M.A., Karmirantzou M., Snel B., Vilanova D., Berger B.,
RA Pessi G., Zwahlen M.-C., Desiere F., Bork P., Delley M.,
RA Fridmore R.D., Arigoni F.;
RT "The genome sequence of Bifidobacterium longum reflects its adaptation
RT to the human gastrointestinal tract."
RL Proc. Natl. Acad. Sci. U.S.A. 99:14422-14427(2002).
CC -!- CATALYTIC ACTIVITY: ATP + L-arginine + tRNA(Arg) = AMP +
CC diphosphate + L-arginyl-tRNA(Arg).
CC -!- SUBUNIT: Monomer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; AB014754; AAN25073.1; -.
CC HAMAP; MF_00123; -.
CC InterPro; IPR001278; Arg_tRNA-synt_1c.
CC InterPro; IPR005148; N.
CC InterPro; IPR008909; tRNA-synt_1d_c.
CC InterPro; IPR001412; tRNA-synt_1i.
CC InterPro; IPR009080; tRNA-synt_1a_bind.
CC Pfam; PF03485; NArg; 1.
CC Pfam; PF00750; tRNA-synt_1d; 1.

Query Match 55.9%; Score 38; DB 1; Length 3503;
Best Local Similarity 60.0%; Pred. No. 1.7e+02;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 DVPYPTSYRY 10
   | : | | | |
DB 149 DUEPYNTQRY 158

RESULT 13
RS1_PROSP
ID RS1_PROSP STANDARD; PRT; 378 AA.
AC P14128;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 30S ribosomal protein S1 (Fragment).
GN RPSA.
OS Providencia sp.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Providencia.
OX NCBI_TaxID=589;
RN [1]
RC MEDLINE=86013618; PubMed=3862933;
RX Schnier J., Paist G.;
RT "Comparative studies on the structural gene for the ribosomal protein
RT S1 in ten bacterial species."
RL Mol. Gen. Genet. 200:476-481(1995).
CC -!- FUNCTION: BINDS MRNA; THUS FACILITATING RECOGNITION OF THE
CC INITIATION POINT. IT IS NEEDED TO TRANSLATE MRNA WITH A SHORT
CC SHINE-DALGARNO (SD) PURINE-RICH SEQUENCE (BY SIMILARITY).
CC -!- SIMILARITY: Belongs to the S1P family of ribosomal proteins.
CC -!- SIMILARITY: Contains 5 S1 motif domains.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; X02828; CAA36596.1; -.
CC HSP; P05055; 1SRO.
CC InterPro; IPR008994; Nucleic acid OB.
CC InterPro; IPR001110; Ribosomal_S1.
CC InterPro; IPR003029; S1.
CC Pfam; PF00575; S1; 5.
CC PRINTS; PR00681; RIBOSOMALS1.
CC SMART; SM00316; S1; 4.
CC PROSITE; PS50126; S1; 5.
CC Ribosomal protein; Repeat; RNA-binding.
FT NON_TER 1 1
FT DOMAIN <1 66 S1 MOTIF 1.
FT DOMAIN 67 155 S1 MOTIF 2.
FT DOMAIN 172 242 S1 MOTIF 3.
FT DOMAIN 229 329 S1 MOTIF 4.
FT DOMAIN 346 378 S1 MOTIF 5.
SQ SEQUENCE 378 AA; 41613 MW; 9052424346D2E904 CRC64;

Query Match 52.9%; Score 36; DB 1; Length 378;
Best Local Similarity 58.3%; Pred. No. 34;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
```

```
DR Pfam; PF05746; tRNA-synt_1d_c; 1.
DR PRINTS; PR01038; TRNASYNTHARG.
DR TIGRFAMs; TIGR00456; args; 1.
DR PROSITE; PS00178; AA tRNA_LIGASE I; 1.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW Complete proteome. 157 "HIGH" REGION
FT SITE 147.
SQ SEQUENCE 620 AA; 67206 MW; F51C3C0C8CD9AB06 CRC64;

Query Match 54.4%; Score 37; DB 1; Length 620;
Best Local Similarity 50.0%; Pred. No. 39;
Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 DVPYPTSYRY 12
   | : | | | |
DB 534 DRQPHKVARYLE 545

RESULT 13
RS1_PROSP
ID RS1_PROSP STANDARD; PRT; 378 AA.
AC P14128;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 30S ribosomal protein S1 (Fragment).
GN RPSA.
OS Providencia sp.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Providencia.
OX NCBI_TaxID=589;
RN [1]
RC MEDLINE=86013618; PubMed=3862933;
RX Schnier J., Paist G.;
RT "Comparative studies on the structural gene for the ribosomal protein
RT S1 in ten bacterial species."
RL Mol. Gen. Genet. 200:476-481(1995).
CC -!- FUNCTION: BINDS MRNA; THUS FACILITATING RECOGNITION OF THE
CC INITIATION POINT. IT IS NEEDED TO TRANSLATE MRNA WITH A SHORT
CC SHINE-DALGARNO (SD) PURINE-RICH SEQUENCE (BY SIMILARITY).
CC -!- SIMILARITY: Belongs to the S1P family of ribosomal proteins.
CC -!- SIMILARITY: Contains 5 S1 motif domains.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; X02828; CAA36596.1; -.
CC HSP; P05055; 1SRO.
CC InterPro; IPR008994; Nucleic acid OB.
CC InterPro; IPR001110; Ribosomal_S1.
CC InterPro; IPR003029; S1.
CC Pfam; PF00575; S1; 5.
CC PRINTS; PR00681; RIBOSOMALS1.
CC SMART; SM00316; S1; 4.
CC PROSITE; PS50126; S1; 5.
CC Ribosomal protein; Repeat; RNA-binding.
FT NON_TER 1 1
FT DOMAIN <1 66 S1 MOTIF 1.
FT DOMAIN 67 155 S1 MOTIF 2.
FT DOMAIN 172 242 S1 MOTIF 3.
FT DOMAIN 229 329 S1 MOTIF 4.
FT DOMAIN 346 378 S1 MOTIF 5.
SQ SEQUENCE 378 AA; 41613 MW; 9052424346D2E904 CRC64;

Query Match 52.9%; Score 36; DB 1; Length 378;
Best Local Similarity 58.3%; Pred. No. 34;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
```

```

SQ      SEQUENCE       396 AA;   42709 MW;   E0FD2A96E9785D81 CRC64;

Query_Match          52.9%;   Score 36;   DB 1;   Length 396;
Best Local Similarity 46.2%;   Pred. No. 36;
Matches               6;   Conservative         4;   Mismatches    3;   Indels        0;   Gaps            0;

Qy      1 DVPRYRTSYLEL 13
|:||||::|||
Db      336 DLRPYAITKNWLDL 348

RESULT 15
YBGG_ECOLI STANDARD; PRT; 415 AA.
ID YBGG_ECOLI STANDARD; PRT; 415 AA.
AC P39455; F77602;
DT 01-FEB-1995 (Rel. 31, Created)
DDT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical protein ybdg.
GN YBGG OR B0577 OR SF0483 OR S0492.
OS Escherichia coli, and
OS Shigella flexneri.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
NCBI TaxID=562, 623,
[1]_
RN [1]
RC SEQUENCE FROM N.A.
RP SPECIES=E.coli; STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shaoy Y.;
RA "The complete genome sequence of Escherichia coli K-12.";
RT Science 277:1453-1474 (1997).
RL [2]
RN [2]
RC SEQUENCE FROM N.A.
RP SPECIES=E.coli; STRAIN=K12 / MG1655;
RX Oshima T., Alba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA Ikenoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA Yano M., Horiiuchi T.;
RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome
corresponding to the 12.7-28.0 min region on the linkage map.";
RL DNA Res. 3:137-155(1996).
RL [4]
RN [4]
RC SEQUENCE OF 1-77 FROM N.A.
RP SPECIES=E.coli; STRAIN=B;
RX MEDLINE=95113294; PubMed=7813889;
RA Michael N.P., Brehm J.K., Anlezark G.M., Minton N.P.;
RT "physical characterisation of the Escherichia coli B gene encoding
nitrosodutase and its over-expression in Escherichia coli K12.";
RL FEBS Microbiol. Lett. 124:195-202(1994).
RL [5]
RN [5]
RC SEQUENCE OF 1-12 FROM N.A.
RP SPECIES=E.coli; STRAIN=K12 / C600;
RX Zerno S., Koike H., Tanokura M., Saigo K.;
RT Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.
RL [6]
RN [6]
RC SEQUENCE FROM N.A.
RP SPECIES=S.flexneri; STRAIN=301 / Serotype 2a;
RX MEDLINE=22272406; PubMed=12384590;
RA Jin O., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,
```

```

1 DVPRTSYL 12
||||| : |||
32 DVPVRTTH 43

RESULT 14
METK PSRAE STANDARD; PRT; 396 AA.
Q91520;
AC 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE S-adenosylmethionine synthetase (EC 2.5.1.6) (Methionine
EN adenosyltransferase) (AdoMet synthetase) (MAT).
METK OR PA0546.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=287;
[1]
RP SEQUENCE FROM N.A.
RN MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey M.J., Brinkman F.S.L., Huftagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltry L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.A., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sater M.H., Hancock R.E.W., Lory S., Olson M.V.;
RA "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen.";
RL Nature 406:959-964 (2000).
CC -! FUNCTION: Catalyzes the formation of S-adenosylmethionine from
CC methionine and ATP. The overall synthetic reaction is composed of
CC two sequential steps, AdoMet formation and the subsequent
CC triolophosphate hydrolysis which occurs prior to release of
CC AdoMet from the enzyme (By similarity).
CC -! CATALYTIC ACTIVITY: ATP + L-methionine + H(2)O = phosphate +
CC diphosphate + S-adenosyl-L-methionine.
CC -! COFACTOR: Binds 2 divalent ions, such as magnesium or cobalt, and
CC 1 potassium ion per subunit (By similarity).
CC -! PATHWAY: Activated methyl cycle.
CC -! SUBUNIT: Homotetramer (By similarity).
CC -! SUBCELLULAR LOCATION: Cytoplasmic.
CC -! SIMILARITY: Belongs to the AdoMet synthetase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE004491; AAC03935.1; -.
CC FIR; H83576; H83576.
CC HSSP; P04384; IMXB.
CC HMAP; MF 00086; -.
CC InterPro; IPR002133; S-AdoMet_synt.
CC Pfam; PF00438; S-AdoMet_synt; 1.
CC Pfam; PF02772; S-AdoMet_synt2; 1.
CC Pfam; PF02773; S-AdoMet_synt3; 1.
CC TIGRfams; TIGR01034; metK; 1.
CC PROSITE; PS00376; ADO MET SYNTHETASE 1; 1.
CC PROSITE; PS00377; ADO MET SYNTHETASE 2; 1.
CC Transferase; One-carbon metabolism; ATP-binding; Magnesium; Potassium;
CC Metal-binding; Complete proteome.
CC NE BIND 261 268 ATP (POTENTIAL).
CC FT METAL 18 18 MAGNESIUM (BY SIMILARITY).
CC FT METAL 44 44 POTASSIUM (BY SIMILARITY).
CC FT METAL 265 265 POTASSIUM (BY SIMILARITY).
CC FT METAL 273 273 MAGNESIUM (BY SIMILARITY).

```

RA Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,  
 RA Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,  
 RA Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,  
 RA Yu J.,  
 RT "Genome sequence of Shigella flexneri 2a: insights into pathogenicity  
 RT through comparison with genomes of Escherichia coli K12 and O157.",  
 RL Nucleic Acids Res. 30:4432-4441 (2002).  
 RN [?]  
 RP SEQUENCE FROM N.A.  
 RP SPECIES=S.flexneri; STRAIN=2457T / ATCC 700930 / Serotype 2a;  
 RX MEDLINE=22590274; PubMed=12704152;  
 RA Wei J., Goldberg M.B., Burland V., Venkatesan M.M., Deng W.,  
 RA Fournier G., Mayhew G.F., Plunkett G. III, Rose D.J., Darling A.,  
 RA Mau B., Perna N.T., Payne S.M., Runyen-Janecky L.J., Zhou S.,  
 RA Schwartz D.C., Blattner F.R.;  
 RT "Complete genome sequence and comparative genomics of Shigella  
 RT flexneri serotype 2a strain 2457T";  
 RL Infect. Immun. 71:2775-2786 (2003).  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AB000163; AAC73678.1; -.  
 DR EMBL; U82598; AAB40775.1; -.  
 DR EMBL; D90700; BAA35217.1; -.  
 DR EMBL; U07860; AAC43265.1; -.  
 DR EMBL; D25414; -: NOT ANNOTATED CDS.  
 DR EMBL; AB015080; AAN42134.1; ALT\_INIT.  
 DR EMBL; AB016979; AAP16008.1; -.  
 DR PIR; G64790; G64790.  
 DR EcoGene; EG12620; ybdG.  
 DR InterPro; IPR006685; MSion\_channel.  
 DR Pfam; PF00924; MS\_channel; 1.  
 KW Hypothetical protein; Transmembrane; Complete proteome.  
 FT TRANSMEM 25 45 POTENTIAL.  
 FT TRANSMEM 68 88 POTENTIAL.  
 FT TRANSMEM 105 125 POTENTIAL.  
 FT TRANSMEM 149 169 POTENTIAL.  
 FT TRANSMEM 174 194 POTENTIAL.  
 FT TRANSMEM 370 390 POTENTIAL.  
 SQ SEQUENCE 415 AA; 46602 MW; C9A8A8A41CDC6F28 CRC64;

Query Match 52.9%; Score 36; DB 1; Length 415;  
 Best Local Similarity 50.0%; Pred. No. 38;  
 Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 2 VRPYRTSRYLEL 13  
 DB 291 LXPYLTSRHQEI 302

Search completed: October 5, 2004, 08:01:38  
 Job time : 2.16358 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2004, 07:34:11, Search time 6.88117 Seconds  
(without alignments)  
596.081 Million cell updates/sec

Title: US-09-805-290A-9

Perfect score: 68

Sequence: 1 DVRPYRTSRYLEL 13

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database:

SPTREMBL 25:\*

1: sp\_archaea:\*

2: sp\_bacteria:\*

3: sp\_fungi:\*

4: sp\_human:\*

5: sp\_invertebrate:\*

6: sp\_mammal:\*

7: sp\_mhc:\*

8: sp\_organelle:\*

9: sp\_phase:\*

10: sp\_plant:\*

11: sp\_rodent:\*

12: sp\_virus:\*

13: sp\_vertebrate:\*

14: sp\_unclassified:\*

15: sp\_rvirus:\*

16: sp\_bacteriap:\*

17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	43	63.2	272	10 Q8LAQ8	Q8LAQ8 arabidopsis
2	42	61.8	462	11 Q8CON9	Q8CON9 mus musculus
3	41	60.3	322	16 Q7WCG8	Q7WCG8 bordetella
4	41	60.3	322	16 Q7W4Z0	Q7W4Z0 bordetella
5	41	60.3	775	10 Q9S1Y9	Q9S1Y9 arabidopsis
6	40	58.8	441	16 Q8X372	Q8X372 escherichia
7	40	58.8	442	16 Q8U9Y6	Q8U9Y6 agrobacteri
8	40	58.8	451	11 Q35390	Q35390 rattus norv
9	40	58.8	581	10 Q9LTU1	Q9LTU1 arabidopsis
10	40	58.8	654	2 Q52001	Q52001 streptococc
11	40	58.8	766	5 P91530	P91530 ciona intes
12	39	57.4	91	13 Q9YHE1	Q9YHE1 brachydanio
13	39	57.4	196	2 Q7WZ84	Q7WZ84 nonomuraea
14	39	57.4	217	5 Q25368	Q25368 loligo opal
15	39	57.4	258	10 Q80652	Q80652 arabidopsis
16	39	57.4	303	10 Q9MLK7	Q9MLK7 arabidopsis

17	39	57.4	323	17 Q8TGZ6	Q8TGZ6 methanopyru
18	39	57.4	393	13 Q803R3	Q803R3 brachydanio
19	39	57.4	558	16 Q87BJ7	Q87BJ7 xylella fas
20	39	57.4	559	16 Q8P8P5	Q8P8P5 xanthomonas
21	39	57.4	561	16 Q8PK77	Q8PK77 xanthomonas
22	39	57.4	569	16 Q9PAQ7	Q9PAQ7 xylella fas
23	39	57.4	980	10 Q84SFS	Q84SFS oryza sativ
24	38	55.9	51	16 Q7UW68	Q7UW68 rhodospirill
25	38	55.9	281	16 Q89AB5	Q89AB5 rhodospirill
26	38	55.9	290	4 Q9H9D2	Q9H9D2 homo sapien
27	38	55.9	346	5 Q9N575	Q9N575 caenorhabdi
28	38	55.9	417	16 Q92RU9	Q92RU9 rhizobium m
29	38	55.9	499	4 Q7Z3V7	Q7Z3V7 homo sapien
30	38	55.9	2870	16 Q97LF2	Q97LF2 clostridium
31	37	54.4	64	16 Q3JUT5	Q3JUT5 neisseria m
32	37	54.4	136	16 Q8PF48	Q8PF48 xanthomonas
33	37	54.4	225	8 Q9T9H4	Q9T9H4 halocynthia
34	37	54.4	254	5 Q9VJY1	Q9VJY1 drosophila
35	37	54.4	254	5 Q8MZ99	Q8MZ99 drosophila
36	37	54.4	264	16 Q82PY4	Q82PY4 streptomyce
37	37	54.4	286	17 Q54583	Q54583 halobacteri
38	37	54.4	371	16 Q83FNO	Q83FNO tropheryma
39	37	54.4	422	17 Q974N2	Q974N2 sulfolobus
40	37	54.4	509	5 Q01792	Q01792 caenorhabdi
41	37	54.4	571	16 Q8Y294	Q8Y294 ralstonia s
42	37	54.4	599	10 P93797	P93797 volvox cart
43	37	54.4	600	5 Q8IMF0	Q8IMF0 drosophila
44	37	54.4	620	16 Q8G4V2	Q8G4V2 bifidobacte
45	37	54.4	868	10 Q9FN94	Q9FN94 arabidopsis

## ALIGNMENTS

### RESULT 1

Q8LAQ8 ID Q8LAQ8 PRELIMINARY; PRT; 272 AA.

AC Q8LAQ8; DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Hypothetical protein.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Haas B.J., Volfovsky N., Town C.D., Troukhan M., Alexandrov N.,  
RA Feldmann K.A., Flavell R.B., White O., Salzberg S.L.;  
RT "Full-length messenger RNA sequences greatly improve genome  
RT annotation.";  
RL Genome Biol. 0:0-0(2002).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Brover V., Troukhan M., Alexandrov N., Lu Y.-P., Flavell R.,  
RA Feldmann K.;  
RT "Full-length cDNA from Arabidopsis thaliana.";  
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AY087671; AAC65208.1; -  
DR InterPro; IPR005162; Retrotrans\_gag.  
DR Pfam; PF03732; Retrotrans\_gag; 1.  
KW Hypothetical protein.  
SQ SEQUENCE 272 AA; 30811 MW; 9C889BD760029D09 CRC64;

Query Match 63.2%; Score 43; DB 10; Length 272;

Best Local Similarity 53.8%; Pred. No. 8.2;

Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 DVRPYRTSRYLEL 13

Db 97 DTHPHRMSRYIQL 109

Query Match 60.3%; Score 41; DB 16; Length 322;  
Best Local Similarity 61.5%; Pred. No. 23;

RA Copenhagen G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,  
 RA Salzberg S.L., Fraser C.M., Venter J.C.;  
 RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis  
 RT thaliana";  
 RL Nature 402:761-768(1999).  
 [2]  
 RP SEQUENCE FROM N.A.  
 RN STRAIN=cv. Columbia;  
 RC Lin X.;  
 RA Submitted (XAR-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AC007020; AAD25679.1; -;  
 DR PIR; B84828; B84828.  
 DR InterPro; IPR001680; WD40.  
 DR Pfam; PF00400; WD40; 4.  
 DR SMART; SMC0320; WD40; 4.  
 DR PROSITE; PS00675; WD\_REPEATS\_1; 1.  
 DR PROSITE; PS00082; WD\_REPEATS\_2; 1.  
 DR PROSITE; PS0294; WD\_REPEATS\_REGION; 2.  
 KW Repeat; WD repeat.  
 SQ SEQUENCE 775 AA; 88051 MW; CBFC6F54B528F4EA CRC64;

Query Match 60.3%; Score 41; DB 10; Length 775;

Best Local Similarity 66.7%; Pred. No. 59; Mismatches 1; Indels 0; Gaps 0;

QY 1 DVRPRTSYLE 12  
 :|||:|  
 DB 429 DLRPNPSCYLE 440

RESULT 6

ID Q8X972 PRELIMINARY; PRT; 441 AA.  
 AC Q8X972;  
 DT 01-MAR-2002 (TREMREL. 20, Created)  
 DT 01-MAR-2002 (TREMREL. 20, Last sequence update)  
 DT 01-OCT-2003 (TREMREL. 25, Last annotation update)  
 DE Hypothetical protein z3943.  
 GN Z3943 OR ECS3511.  
 OS Escherichia coli O157:H7.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Escherichia.  
 CX NCBI\_TaxID=83334;  
 RN [1]

RP SEQUENCE FROM N.A.  
 RC STRAIN=O157:H7 / EDL933 / ATCC 700927;  
 RX MEDLINE=21074935; PubMed=11206551;  
 RA Perna N.T., Plunkett G., III, Burland V., Mau B., Glasner J.D.,  
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,  
 RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,  
 RA Grobeck E.J., Davis N.W., Lam A., Dimalanta E.T., Potamousis K.,  
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,  
 RA Welch R.A., Blattner F.R.;  
 RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7";  
 RL Nature 409:529-533(2001).  
 [2]

RP SEQUENCE FROM N.A.  
 RC STRAIN=O157:H7 / RIMD 0509952;  
 RX MEDLINE=21156231; PubMed=11258796;  
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,  
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,  
 RA Iida T., Takami H., Honda T., Sasaki C., Ogasawara N., Yasunaga T.,  
 RA Kuwara S., Saiba T., Hattori M., Shinagawa H.;  
 RT "Complete genome sequence of enterohemorrhagic Escherichia coli  
 RT O157:H7 and genomic comparison with a laboratory strain K-12";  
 RL DNA Res. 8:11-22(2001).  
 DR EMBL; AE005493; AAG57756.1; -;  
 DR EMBL; AP002562; BAB36934.1; ALT\_INIT.  
 DR PIR; G91067; G91067.  
 DR PIR; H85911; H85911.  
 DR GO; GO:0003677; F:DNA binding; IEA.  
 DR GO; GO:0015074; P:DNA integration; IEA.  
 DR GO; GO:0006310; P:DNA recombination; IEA.

DR InterPro; IPR002104; Phage integrase.  
 DR Pfam; PF00589; Phage integrase; 1.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 441 AA; 51923 MW; 788E55C8CD670D CRC64;

Query Match 58.8%; Score 40; DB 16; Length 441;  
 Best Local Similarity 61.5%; Pred. No. 50;  
 Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 DVRPRTSYLE 13  
 :|||:|  
 DB 321 DVRSERTNSYIEL 333

RESULT 7

ID Q8U8Y6 PRELIMINARY; PRT; 442 AA.  
 AC Q8U8Y6;  
 DT 01-JUN-2002 (TREMREL. 21, Created)  
 DT 01-JUN-2002 (TREMREL. 21, Last sequence update)  
 DT 01-OCT-2003 (TREMREL. 25, Last annotation update)  
 DE Oxidoreductase.  
 GN ATU3952 OR AGR\_L1801.  
 OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
 OC Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.  
 CX NCBI\_TaxID=176299;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21608550; PubMed=11743193;  
 RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,  
 RA Okura V.K., Zhou Y., Chen L., Wood G.P., Almeida N.F. Jr., Woo L.,  
 RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Bovee D. Sr.,  
 RA Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,  
 RA Kutavain T., Levy R., Li M.-J., McClelland E., Palmieri A.,  
 RA Raymond C., Rouse G., Saenphimmachak C., Wu Z., Romero P., Gordon D.,  
 RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,  
 RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,  
 RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,  
 RA Nester E.W.;  
 RT "The genome of the natural genetic engineer Agrobacterium tumefaciens  
 RT C58";  
 RL Science 294:2317-2323(2001).  
 [2]

RP SEQUENCE FROM N.A.  
 RX MEDLINE=21608551; PubMed=11743194;  
 RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,  
 RA Qurollo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,  
 RA Houmel K., Gordon J., Vaudin M., Iatichouk O., Epp A., Liu F.,  
 RA Wollam C., Allinger M., Doughty D., Scott C., Lappas C., Markelz B.,  
 RA Flanagan C., Crowell C., Gurson J., Lomo C., Sear C., Strub G.,  
 RA Cielo C., Slater S.;  
 RT "Genome sequence of the plant pathogen and biotechnology agent  
 RT Agrobacterium tumefaciens C58";  
 RL Science 294:2323-2328(2001).  
 DR EMBL; AE009126; AAL44754.1; -;  
 DR EMBL; AE008288; AAK89474.1; -;  
 DR PIR; AD3042; AD3042.  
 DR PIR; H98243; H98243.  
 DR GO; GO:0016491; F:oxidoreductase activity; IEA.  
 DR GO; GO:0006118; P:electron transport; IEA.  
 DR InterPro; IPR006076; FAD\_oxred.  
 DR InterPro; IPR000205; NAD\_BS.  
 DR Pfam; PF01266; DAO; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 442 AA; 48309 MW; B074DB4BCFFB62D CRC64;

Query Match 58.8%; Score 40; DB 16; Length 442;  
 Best Local Similarity 58.3%; Pred. No. 50;  
 Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 DVRPRTSYLE 12  
 :|||:|

Db 421 DRRPYRSRFFD 432

# RESULT 8

O35390 PRELIMINARY; PRT; 451 AA.  
 AC O35390;  
 DT 01-JAN-1998 (TREMELrel. 05, Created)  
 DT 01-JAN-1998 (TREMELrel. 05, Last sequence update)  
 DT 01-NOV-1998 (TREMELrel. 08, Last annotation update)  
 DE Endo-alpha-D-mannosidase.  
 GN ENMAN.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]\_TaxID=10116;  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Sprague-Dawley;  
 RX MEDLINE=98030622; PubMed=9361017;  
 RA Spiro M.J., Bhoyroo V.D., Spiro R.G.;  
 RT "Molecular cloning and expression of rat liver endo-alpha-mannosidase,  
 an N-linked oligosaccharide processing enzyme.";  
 RL J. Biol. Chem. 272:29356-29363(1997).  
 DR EMBL: AF023657; AAB86925.1; -. 5A987BFD07585066 CRC64;  
 SQ SEQUENCE 451 AA; 51671 MW; 5A987BFD07585066 CRC64;

Query Match 58.8%; Score 40; DB 11; Length 451;

Best Local Similarity 53.8%; Pred. No. 51;

Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Oy 1 DRRPYRSRYLEL 13

Db 415 DRRPKPSLYLEI 427

# RESULT 9

O9LTU1 PRELIMINARY; PRT; 581 AA.  
 AC O9LTU1;  
 DT 01-OCT-2000 (TREMELrel. 15, Created)  
 DT 01-OCT-2000 (TREMELrel. 15, Last sequence update)  
 DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)  
 DE Replication protein A1-like.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsia.  
 OX NCBI\_TaxID=3702;  
 RN [1]\_TaxID=3702;  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Columbia;  
 RA Sato S., Nakamura Y., Kaneko T., Kato T., Asamizu E., Tabata S.;  
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.  
 RN [2]

SEQUENCE FROM N.A.  
 AC STRAIN=Columbia;  
 RX MEDLINE=20277480; PubMed=10819329;  
 RA Nakamura Y.;  
 RT "Structural analysis of Arabidopsis thaliana chromosome 3. I. Sequence  
 features of the regions of 4,504,864 bp covered by sixty pl and TAC  
 clones.";  
 RT DNA Res. 7:131-135(2000).  
 RL EMBL: AB024034; BAB02796.1; -.  
 DR InterPro: IPR003871; DUF223.  
 DR InterPro: IPR008934; Nucleic\_acid\_OB.  
 DR Pfam: PF02721; DUF223; 1.  
 SQ SEQUENCE 581 AA; 64867 MW; 5BCB4DB5B0E006DC CRC64;

Query Match 58.8%; Score 40; DB 10; Length 581;

Best Local Similarity 87.5%; Pred. No. 67;

Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 1 DRRPYRPTS 8  
 Db 10 DRRPYKTS 17

# RESULT 10

O52001 PRELIMINARY; PRT; 654 AA.  
 AC O52001;  
 DT 01-NOV-1996 (TREMELrel. 01, Created)  
 DT 01-NOV-1996 (TREMELrel. 01, Last sequence update)  
 DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)  
 DE NICKASE.  
 OS Streptococcus sp., and  
 OS Enterococcus faecalis (Streptococcus faecalis).  
 OG Plasmid pIP501, and Plasmid PRE25.  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=1306, 1351;  
 RN [1]\_TaxID=1306, 1351;  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=Streptococcus sp.; STRAIN=PVA1702; PLASMID=PIP501;  
 RX MEDLINE=95362640; PubMed=7635806;  
 RA Wang A., Macrina F.L.;  
 RT "Streptococcal plasmid pIP501 has a functional *oriT* site.";  
 RL J. Bacteriol. 177:4199-4206(1995).  
 RN [2]

# SEQUENCE FROM N.A.

RC SPECIES=E.faecalis; STRAIN=RE25; PLASMID=PRE25;

RX PubMed=11735367;

RA Schwarz F.V., Perreten V., Teuber M.;

RT "Sequence of the 50-kb conjugative multiresistance plasmid PRE25 from

Enterococcus faecalis RE25.";

RL Plasmid 46:170-187(2001).

DR EMBL: L39769; AAA99466.1; -.

DR EMBL: X92945; CAC39179.1; -.

DR GO: 0046821; C:extrachromosomal DNA; IEA.

DR GO: 0009291; P:unidirectional conjugation; IEA.

DR InterPro: IPR005053; MobA\_MobL.

DR Pfam: PF03389; MobA\_MobL; 1.

KW Plasmid.

SQ SEQUENCE 654 AA; 76493 MW; C3E5B8BB4468600 CRC64;

Query Match 59.8%; Score 40; DB 2; Length 654;

Best Local Similarity 60.0%; Pred. No. 76;

Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Oy 3 RPYRTSYLYE 12

Db 418 KPFQTSRYLD 427

# RESULT 11

P91590 PRELIMINARY; PRT; 766 AA.  
 AC P91590;  
 DT 01-MAY-1997 (TREMELrel. 03, Created)  
 DT 01-NOV-1998 (TREMELrel. 08, Last sequence update)  
 DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)  
 DE Transglutiminase.  
 OS Ciona intestinalis.  
 OC Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;  
 OC Phlebobranchia; Cionidae; Ciona.  
 OX NCBI\_TaxID=7719;  
 RN [1]\_TaxID=7719;  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97330926; PubMed=9187361;  
 RA Cariello L., Ristoratore F., Zanetti L.;  
 RT "A new transglutaminase-like from the ascidian *Ciona intestinalis*.";  
 RL FEBS Lett. 408:171-176(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Cariello L.;



RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.

DR ENBL; Y10212; CAA71263.1; --

DR HSP; P00488; LGSU.

DR InterPro; IPR001102; GluttransfG.

DR InterPro; IPR007110; Ig-like.

DR InterPro; IPR008958; Transglut C.

DR InterPro; IPR002931; Trnsglutase like.

DR Pfam; PF00927; Transglutamin C; 1.

DR Pfam; PF00868; Transglutamin N; 1.

DR Pfam; PF01841; Transglut\_core; 1.

DR SMART; SM00460; TGC; 1.

DR SEQUENCE 766 AA; 87194 MW; 7F90D921D87FDDID CRC64;

Query Match 58.8%; Score 40; DB 5; Length 766;

Best Local Similarity 77.8%; Pred. No. 89;

Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 DVPRYRTSR 9

Db 732 EVRPYRSTR 740

RESULT 12

QYVHEI

ID QYVHEI PRELIMINARY; PRT; 91 AA.

AC QYVHEI

DT 01-MAY-1999 (TrEMBLrel. 10, Created)

DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)

DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)

DE Protein tyrosine phosphatase 1B (EC 3.1.3.48) (Fragment).

OS Brachydanio rerio (Zebrafish) (Danio rerio).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Cyprinidae; Danio.

OC NCBI\_TaxID=7955;

[1]

SEQUENCE FROM N.A.

Kral L.G., Cornwell J.R. Jr., Barron C.S. III, Marty S.;

"Intracellular protein tyrosine phosphatases expressed in zebrafish

embryos: New members of PTP-IB and PTP-MEGL related protein tyrosine

phosphatase sub-families."

RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.

ENBL; AF097483; AAC97944.1; --

HSP; P18031; 1PTY.

DR GO; GO:0016787; F:hydrolase activity; IEA.

DR GO; GO:0004727; F:prenylated protein tyrosine phosphatase act. .; IEA.

DR GO; GO:0006470; P:protein amino acid dephosphorylation; IEA.

DR InterPro; IPR000242; Tyr\_PP.

DR Pfam; PF00102; Y phosphatase; 1.

DR PRINTS; PR00700; PRTYPHPTASE.

DR PROSITE; PS50055; TYR\_PHOSPHATASE\_PTP; 1.

KW Hydrolase.

FT NON\_TER 1 1

FT NON\_TER 91 91

SQ SEQUENCE 91 AA; 10669 MW; E901B54F19E930B8 CRC64;

Query Match 57.4%; Score 39; DB 13; Length 91;

Best Local Similarity 61.5%; Pred. No. 15;

Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 DVPRYRTSRVLEL 13

Db 25 DVKSYYTTRVLEL 37

RESULT 13

ID Q7WZ84

AC Q7WZ84 PRELIMINARY; PRT; 196 AA.

DT 01-OCT-2003 (TrEMBLrel. 25, Created)

DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Putative vanY-type carboxypeptidase.

GN DBV7.

OS Nonomuraea sp. ATCC 39727.

OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;

OC Streptosporangineae; Streptosporangiaceae; Nonomuraea.

OX NCBI\_TaxID=93944;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-ATCC 39727;

RA Sosio M., Stinchi S., Beltrametti F., Lazzarini A., Donadio S.;

"The gene cluster for the biosynthesis of the glycopeptide antibiotic

A40926 by Nonomuraea sp.;"

RT Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.

DR EMBL; AJ561198; CAD91202.1; --

KW Acyltransferase; Carboxypeptidase; Glycosyltransferase; Monooxygenase.

SQ SEQUENCE 196 AA; 22157 MW; 003AA44D19127982 CRC64;

Query Match 57.4%; Score 39; DB 2; Length 196;

Best Local Similarity 58.3%; Pred. No. 33;

Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 DVPRYRTSRYLE 12

Db 137 DVPRHEGARWLE 148

RESULT 14

Q25368

ID Q25368 PRELIMINARY; PRT; 217 AA.

AC Q25368;

DT 01-NOV-1996 (TrEMBLrel. 01, Created)

DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)

DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)

DE S-crystallin.

OC Loligo opalescens (California market squid).

OC Eukaryota; Metazoa; Mollusca; Cephalopoda; Coleoidea; Neocoleoidea;

OC Decapodiformes; Loliginidae; Loligo.

OX NCBI\_TaxID=31211;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Lens;

RX MEDLINE=96139039; PubMed=8587103;

RA Tomarev S.I., Chung S., Piatigorsky J.;

"Glutathione S-transferase and S-crystallins of cephalopods: evolution

from active enzyme to lens-refractive proteins.;"

RL J. Mol. Evol. 41:1048-1056(1995).

DR EMBL; U19300; AAA97551.1; --

DR HSP; P46088; 2GSQ.

DR InterPro; IPR004046; GST Cterm.

DR InterPro; IPR004045; GST Nterm.

DR InterPro; IPR003083; S-crystallin.

DR Pfam; PF00043; GST\_C; 1.

DR Pfam; PF02798; GST\_N; 1.

DR PRINTS; PR01269; SCRYSTALLIN.

SQ SEQUENCE 217 AA; 26113 MW; 66D4D97D9B3B6F8A CRC64;

Query Match 57.4%; Score 39; DB 5; Length 217;

Best Local Similarity 50.0%; Pred. No. 36;

Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 DVPRYRTSRYLE 12

Db 125 DLNPMQTRYME 136

RESULT 15

O80652

ID O80652 PRELIMINARY; PRT; 258 AA.

AC O80652;

DT 01-NOV-1998 (TrEMBLrel. 08, Created)

DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)

DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)

DE TL4N5.4 protein.

GN TL4N5.4.

OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsiis.  
 OX NCBI\_TaxID=3702;

RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Columbia;  
 RA Federspiel N.A., Palm C.J., Conway A.B., Kurtz D.B., Conway A.R.,  
 RA Au M., Araujo R., Buehler E., Dewar K., Peng J., Kim C., Li Y.,  
 RA Oji O., Osborne B.I., Shinn P., Sun H., Toriumi M., Vysotskaia V.S.,  
 RA Yu G., Ecker J., Theologis A., Davis R.W.;  
 RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AC004260; AAC34353.1; -.  
 DR FIR; T00447; T00447.  
 DR InterPro; IPR006502; DUF\_plant\_1615.  
 DR Pfam; PF04720; DUF506; 1.  
 DR TIGRFAMs; TIGR01615; A.thal\_3542; 1.  
 SQ SEQUENCE 258 AA; 29578 MW; FBC341434DFA6380 CRC64;

Query Match 57.4%; Score 39; DB 10; Length 258;  
 Best Local Similarity 66.7%; Pred. No. 44;  
 Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 4 PYRTSRYLE 12  
 DB 199 PWRTRRYMQ 207

Search completed: October 5, 2004, 08:13:13  
 Job time : 8.8817 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2004, 07:04:10 ; Search time 9.89043 Seconds  
(without alignments)  
371.381 Million cell updates/sec

Title: US-09-805-290A-9

Perfect score: 68

Sequence: 1 DVRPYRTSRVLEL 13

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_29Jan04.\*

1: geneseqp1900s.\*

2: geneseqp1900s.\*

3: geneseqp2000s.\*

4: geneseqp2001s.\*

5: geneseqp2002s.\*

6: geneseqp2003as.\*

7: geneseqp2003bs.\*

8: geneseqp2004s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	68	100.0	13	4	AAE10541 Llama spe
2	68	100.0	130	4	AAE10554 HPL inhib
3	66	97.1	13	4	AAE10542 Llama spe
4	66	97.1	130	4	AAE10559 HPL inhib
5	65	95.6	13	4	AAE10540 Llama spe
6	65	95.6	130	4	AAE10552 HPL inhib
7	64	94.1	13	4	AAE10543 Llama spe
8	43	63.2	252	3	AAE10543 Llama spe
9	43	63.2	272	3	AAE10543 Llama spe
10	43	63.2	281	3	AAE10543 Llama spe
11	41	60.3	90	6	ABP75655 Human sec
12	40	58.8	142	3	AAE105097 Human sec
13	40	58.8	437	7	AAE101045 Enterchae
14	40	58.8	451	7	AAE101045 Enterchae
15	39	57.4	57	5	ABP35493 Human ORF
16	39	57.4	208	3	AAE10529 Arabidops
17	39	57.4	238	3	AAE10528 Arabidops
18	39	57.4	258	3	AAE10527 Arabidops
19	39	57.4	509	5	ABP90986 Herbicida
20	39	57.4	558	6	ABU33563 Protein e
21	39	57.4	727	6	ABU27962 Protein e
22	39	57.4	944	6	ABU5234 Protein e
23	39	57.4	961	6	ABU32202 Protein e
24	38	55.9	52	4	AAU57763 Propionib
25	38	55.9	52	6	ABM54282 Propionib

26	38	55.9	56	4	AAE04256 Human gen
27	38	55.9	103	7	ADC89474 Ribosomal
28	38	55.9	143	3	AAE45098 Human sec
29	38	55.9	229	5	AAE47969 Human end
30	38	55.9	290	4	AAE95204 Human pro
31	38	55.9	290	7	AAE54768 Human pro
32	38	55.9	378	6	AAE79783 Carbohydr
33	38	55.9	480	6	AAE08478 Alloicoc
34	38	55.9	653	2	AAE54870 Super hea
35	38	55.9	1718	5	AAE26420 Human tra
36	38	55.9	3503	4	AAE66499 Drosophil
37	37	54.4	58	2	AAE87540 Peptide d
38	37	54.4	79	4	AAE24628 Novel hum
39	37	54.4	91	4	AAE62932 Propionib
40	37	54.4	91	6	ABM59451 Propionib
41	37	54.4	250	3	AAE73330 HTRM clon
42	37	54.4	254	4	ABE70887 Drosophil
43	37	54.4	391	7	ADC94988 E. faeciu
44	37	54.4	566	2	AAE87531 An antitu
45	37	54.4	620	5	ABP66018 Bifidobac

## ALIGNMENTS

RESULT 1

AAE10541

ID AAE10541 standard; peptide; 13 AA.

XX AAE10541;

XX AAE10541;

DT 10-DEC-2001 (first entry)

DE Llama species antibody VH CDR3 #10.

XX Llama antibody; camelid; anorectic; heavy chain variable domain; VH;

KW human dietary enzyme inhibitor; medicament; human pancreatic lipase; HPL;

KW food; human gastric lipase; HGL; cosmetic control; body weight;

KW complementarity determining region 3; CDR3.

XX Lama sp.

XX EP1134231-A1.

XX 19-SEP-2001.

XX 20-FEB-2001; 2001EP-00200703.

XX 14-MAR-2000; 2000EP-00200930.

XX (UNIL ) UNILEVER NV.

XX (UNIL ) UNILEVER PLC.

PI Bezemer S, Van De Burg M, De Haard JJW, Tareilus E;

XX WPI; 2001-572718/65.

PT New antibody or its fragments for inhibiting human dietary enzymes,

PT useful for cosmetic control of body weight of human beings, comprises

PT heavy chain variable domain derived from immunoglobulin naturally devoid

XX of light chains.

PS Disclosure; Page 17; 37pp; English.

CC The patent discloses antibodies or their fragments comprising a heavy

CC chain variable domain (VHH) derived from an immunoglobulin naturally

CC devoid of light chains specific for inhibiting human dietary enzymes. The

CC antibodies of the invention are useful for the preparation of medicaments

CC or food for inhibiting the activity of one or more human dietary enzymes

CC especially human pancreatic lipase (HPL) or human gastric lipase (HGL)

CC which are useful for the cosmetic control of body weight of human beings.

CC The present peptide sequence is a complementarity determining region 3

CC (CDR3) of llama species (camelid) antibody VH region

XX SQ Sequence 13 AA; Query Match 100.0%; Score 68; DB 4; Length 13;  
 Best Local Similarity 100.0%; Pred. No. 5.3e-06; Mismatches 0; Indels 0; Gaps 0;  
 Matches 13; Conservative 0; Indels 0; Gaps 0;

QY 1 DVPRPRTSYLEL 13  
 DB 1 DVPRPRTSYLEL 13  
 |||||

RESULT 2  
 AAE10554  
 ID AAE10554 standard; peptide; 130 AA.  
 XX AC AAE10554;  
 XX DT 10-DEC-2001 (first entry)  
 XX DE HPL inhibiting VHH fragment, HPL #14 from llama species.  
 XX KW Llama antibody; camelid; anorectic; heavy chain variable domain; VHH;  
 KW human dietary enzyme inhibitor; medicament; human pancreatic lipase; HPL;  
 KW food; human gastric lipase; HGL; cosmetic control; body weight.  
 XX OS Lama sp.  
 XX PH Key Location/Qualifiers  
 FT Region 31..35  
 FT /label= CDR1  
 FT /note= "Complementarity determining region 1"  
 FT Region 50..64  
 FT /label= CDR2  
 FT /note= "Complementarity determining region 2"  
 FT Region 98..110  
 FT /label= CDR3  
 FT /note= "Complementarity determining region 3"  
 XX EP1134231-Al.  
 XX 19-SEP-2001.  
 XX 20-FEB-2001; 2001EP-00200703.  
 XX 14-MAR-2000; 2000EP-00200930.  
 XX (UNIL ) UNILEVER NV.  
 XX (UNIL ) UNILEVER PLC.  
 XX Bezemer S, Van De Burg M, De Haard JJW, Tareilus E;  
 WPI; 2001-572718/65.  
 XX New antibody or its fragments for inhibiting human dietary enzymes,  
 useful for cosmetic control of body weight of human beings, comprises  
 heavy chain variable domain derived from immunoglobulin naturally devoid  
 of light chains.  
 XX Example 2; Page 9; 37pp; English.  
 XX The patent discloses antibodies or their fragments comprising a heavy  
 chain variable domain (VHH) derived from an immunoglobulin naturally  
 devoid of light chains specific for inhibiting human dietary enzymes. The  
 antibodies of the invention are useful for the preparation of medicaments  
 or food for inhibiting the activity of one or more human dietary enzymes  
 especially human pancreatic lipase (HPL) or human gastric lipase (HGL)  
 which are useful for the cosmetic control of body weight of human beings.  
 The present peptide sequence is HPL inhibiting VHH fragment, HPL #14 from  
 llama (camelid) species  
 XX Sequence 130 AA;

Query Match 100.0%; Score 68; DB 4; Length 130;  
 Best Local Similarity 100.0%; Pred. No. 6.8e-05; Mismatches 0; Indels 0; Gaps 0;  
 Matches 13; Conservative 0; Indels 0; Gaps 0;

QY 1 DVPRPRTSYLEL 13  
 DB 98 DVPRPRTSYLEL 110  
 |||||

RESULT 3  
 AAE10542  
 ID AAE10542 standard; peptide; 13 AA.  
 XX AC AAE10542;  
 XX DT 10-DEC-2001 (first entry)  
 XX DE Llama species antibody VHH CDR3 #11.  
 XX KW Llama antibody; camelid; anorectic; heavy chain variable domain; VHH;  
 KW human dietary enzyme inhibitor; medicament; human pancreatic lipase; HPL;  
 KW food; human gastric lipase; HGL; cosmetic control; body weight;  
 KW complementarity determining region 3; CDR3.  
 XX OS Lama sp.  
 XX PH EP1134231-Al.  
 XX 19-SEP-2001.  
 XX 20-FEB-2001; 2001EP-00200703.  
 XX 14-MAR-2000; 2000EP-00200930.  
 XX (UNIL ) UNILEVER NV.  
 XX (UNIL ) UNILEVER PLC.  
 XX Bezemer S, Van De Burg M, De Haard JJW, Tareilus E;  
 WPI; 2001-572718/65.  
 XX New antibody or its fragments for inhibiting human dietary enzymes,  
 useful for cosmetic control of body weight of human beings, comprises  
 heavy chain variable domain derived from immunoglobulin naturally devoid  
 of light chains.  
 XX Disclosure; Page 17; 37pp; English.  
 XX The patent discloses antibodies or their fragments comprising a heavy  
 chain variable domain (VHH) derived from an immunoglobulin naturally  
 devoid of light chains specific for inhibiting human dietary enzymes. The  
 antibodies of the invention are useful for the preparation of medicaments  
 or food for inhibiting the activity of one or more human dietary enzymes  
 especially human pancreatic lipase (HPL) or human gastric lipase (HGL)  
 which are useful for the cosmetic control of body weight of human beings.  
 The present peptide sequence is a complementarity determining region 3  
 (CDR3) of llama species (camelid) antibody VHH region  
 XX Sequence 13 AA;

Query Match 97.1%; Score 66; DB 4; Length 13;  
 Best Local Similarity 92.3%; Pred. No. 1.3e-05; Mismatches 1; Indels 0; Gaps 0;  
 Matches 12; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DVPRPRTSYLEL 13  
 DB 1 DVPRPRTSYLEI 13  
 |||||

RESULT 4  
 AAE10559  
 ID AAE10559 standard; peptide; 130 AA.  
 XX

```

AC  AAE10559;
XX
XX  10-DEC-2001 (first entry)
XX
XX  HPL inhibiting VHH fragment, HPL #30 from llama species.
XX
XX  Llama antibody; camelid; anorectic; heavy chain variable domain; VHH;
XX  human dietary enzyme inhibitor; medicament; human pancreatic lipase; HPL;
XX  food; human gastric lipase; HGL; cosmetic control; body weight.
XX
XX  Lama sp.
XX
XX  Location/Qualifiers
XX  Key
XX  Region
XX  31..35
XX  /label= CDR1
XX  /note= "Complementarity determining region 1"
XX  Region
XX  49..64
XX  /label= CDR2
XX  /note= "Complementarity determining region 2"
XX  Region
XX  98..110
XX  /label= CDR3
XX  /note= "Complementarity determining region 3"
XX
XX  EPI134231-A1.
XX
XX  19-SEP-2001.
XX
XX  20-FEB-2001; 2001EP-00200703.
XX
XX  14-MAR-2000; 2000EP-00200930.
XX
XX  (UNIL ) UNILEVER NV.
XX  (UNIL ) UNILEVER PLC.
XX
XX  Bezemer S, Van De Burg M, De Haard JJW, Tareilus E;
XX  WPI; 2001-572718/65.
XX
XX  New antibody or its fragments for inhibiting human dietary enzymes,
XX  useful for cosmetic control of body weight of human beings, comprises
XX  heavy chain variable domain derived from immunoglobulin naturally devoid
XX  of light chains.
XX
XX  Example 2; Page 10; 37pp; English.
XX
XX  The patent discloses antibodies or their fragments comprising a heavy
XX  chain variable domain (VHH) derived from an immunoglobulin naturally
XX  devoid of light chains specific for inhibiting human dietary enzymes. The
XX  antibodies of the invention are useful for the preparation of medicaments
XX  or food for inhibiting the activity of one or more human dietary enzymes
XX  especially human pancreatic lipase (HPL) or human gastric lipase (HGL)
XX  which are useful for the cosmetic control of body weight of human beings.
XX  The present peptide sequence is HPL inhibiting VHH fragment, HPL #30 from
XX  llama (camelid) species
XX
XX  Sequence 130 AA;
XX
XX  Query Match 97.1%; Score 66; DB 4; Length 130;
XX  Best Local Similarity 92.3%; Pred. No. 0.00016;
XX  Matches 12; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
XX
XX  1 DVPRPRTSRYLEL 13
XX  |||||
XX  98 DVPRPRTSRYLEI 110
XX
XX  RESULT 5
XX  AAE10540
XX  ID AAE10540 standard; peptide; 13 AA.
XX
XX  AAE10540;
XX
XX  10-DEC-2001 (first entry)
XX

```

```

XX  Llama species antibody VHH CDR3 #9.
XX
XX  Llama antibody; camelid; anorectic; heavy chain variable domain; VHH;
XX  human dietary enzyme inhibitor; medicament; human pancreatic lipase; HPL;
XX  food; human gastric lipase; HGL; cosmetic control; body weight;
XX  complementarity determining region 3; CDR3.
XX
XX  Lama sp.
XX
XX  EPI134231-A1.
XX
XX  19-SEP-2001.
XX
XX  20-FEB-2001; 2001EP-00200703.
XX
XX  14-MAR-2000; 2000EP-00200930.
XX
XX  (UNIL ) UNILEVER NV.
XX  (UNIL ) UNILEVER PLC.
XX
XX  Bezemer S, Van De Burg M, De Haard JJW, Tareilus E;
XX  WPI; 2001-572718/65.
XX
XX  New antibody or its fragments for inhibiting human dietary enzymes,
XX  useful for cosmetic control of body weight of human beings, comprises
XX  heavy chain variable domain derived from immunoglobulin naturally devoid
XX  of light chains.
XX
XX  Disclosure; Page 17; 37pp; English.
XX
XX  The patent discloses antibodies or their fragments comprising a heavy
XX  chain variable domain (VHH) derived from an immunoglobulin naturally
XX  devoid of light chains specific for inhibiting human dietary enzymes. The
XX  antibodies of the invention are useful for the preparation of medicaments
XX  or food for inhibiting the activity of one or more human dietary enzymes
XX  especially human pancreatic lipase (HPL) or human gastric lipase (HGL)
XX  which are useful for the cosmetic control of body weight of human beings.
XX  The present peptide sequence is a complementarity determining region 3
XX  (CDR3) of llama species (camelid) antibody VHH region
XX
XX  Sequence 13 AA;
XX
XX  Query Match 95.6%; Score 65; DB 4; Length 13;
XX  Best Local Similarity 92.3%; Pred. No. 2e-05;
XX  Matches 12; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
XX
XX  1 DVPRPRTSRYLEL 13
XX  |||||
XX  1 DVPRPRTSRYLEV 13
XX
XX  RESULT 6
XX  AAE10552
XX  ID AAE10552 standard; peptide; 130 AA.
XX
XX  AAE10552;
XX
XX  10-DEC-2001 (first entry)
XX
XX  HPL inhibiting VHH fragment, HPL #12 from llama species.
XX
XX  Llama antibody; camelid; anorectic; heavy chain variable domain; VHH;
XX  human dietary enzyme inhibitor; medicament; human pancreatic lipase; HPL;
XX  food; human gastric lipase; HGL; cosmetic control; body weight.
XX
XX  Lama sp.
XX
XX  Key
XX  Region
XX  31..35
XX  /label= CDR1
XX  /note= "Complementarity determining region 1"
XX

```



PR	19-APR-1999;	99US-0130077P.	PR	19-JUL-1999;	99US-0144335P.
PR	21-APR-1999;	99US-0130449P.	PR	20-JUL-1999;	99US-0144352P.
PR	23-APR-1999;	99US-0130510P.	PR	20-JUL-1999;	99US-0144632P.
PR	23-APR-1999;	99US-0130891P.	PR	21-JUL-1999;	99US-0144884P.
PR	28-APR-1999;	99US-0131449P.	PR	21-JUL-1999;	99US-0144814P.
PR	30-APR-1999;	99US-0132048P.	PR	21-JUL-1999;	99US-0145086P.
PR	30-APR-1999;	99US-0132407P.	PR	21-JUL-1999;	99US-0145088P.
PR	04-MAY-1999;	99US-0132484P.	PR	22-JUL-1999;	99US-0145085P.
PR	05-MAY-1999;	99US-0132485P.	PR	22-JUL-1999;	99US-0145087P.
PR	06-MAY-1999;	99US-0132486P.	PR	22-JUL-1999;	99US-0145089P.
PR	07-MAY-1999;	99US-0132487P.	PR	23-JUL-1999;	99US-0145192P.
PR	07-MAY-1999;	99US-0132863P.	PR	23-JUL-1999;	99US-0145145P.
PR	11-MAY-1999;	99US-0134256P.	PR	23-JUL-1999;	99US-0145218P.
PR	14-MAY-1999;	99US-0134218P.	PR	23-JUL-1999;	99US-0145224P.
PR	14-MAY-1999;	99US-0134219P.	PR	26-JUL-1999;	99US-0145276P.
PR	14-MAY-1999;	99US-0134221P.	PR	27-JUL-1999;	99US-0145913P.
PR	14-MAY-1999;	99US-0134370P.	PR	27-JUL-1999;	99US-0145918P.
PR	18-MAY-1999;	99US-0134768P.	PR	27-JUL-1999;	99US-0145919P.
PR	19-MAY-1999;	99US-0134941P.	PR	28-JUL-1999;	99US-0145951P.
PR	20-MAY-1999;	99US-0135124P.	PR	02-AUG-1999;	99US-0146386P.
PR	21-MAY-1999;	99US-0135353P.	PR	02-AUG-1999;	99US-0146389P.
PR	24-MAY-1999;	99US-0135629P.	PR	02-AUG-1999;	99US-0146389P.
PR	25-MAY-1999;	99US-0136021P.	PR	03-AUG-1999;	99US-0147038P.
PR	27-MAY-1999;	99US-0136392P.	PR	04-AUG-1999;	99US-0147204P.
PR	28-MAY-1999;	99US-0136782P.	PR	04-AUG-1999;	99US-0147302P.
PR	01-JUN-1999;	99US-0137222P.	PR	05-AUG-1999;	99US-0147192P.
PR	03-JUN-1999;	99US-0137528P.	PR	05-AUG-1999;	99US-0147260P.
PR	04-JUN-1999;	99US-0137502P.	PR	06-AUG-1999;	99US-0147303P.
PR	07-JUN-1999;	99US-0137724P.	PR	06-AUG-1999;	99US-0147416P.
PR	08-JUN-1999;	99US-0138094P.	PR	09-AUG-1999;	99US-0147493P.
PR	10-JUN-1999;	99US-0138540P.	PR	09-AUG-1999;	99US-0147935P.
PR	10-JUN-1999;	99US-0138847P.	PR	10-AUG-1999;	99US-0148171P.
PR	14-JUN-1999;	99US-0139119P.	PR	11-AUG-1999;	99US-0148319P.
PR	16-JUN-1999;	99US-0139452P.	PR	12-AUG-1999;	99US-0148341P.
PR	17-JUN-1999;	99US-0139453P.	PR	12-AUG-1999;	99US-0148565P.
PR	17-JUN-1999;	99US-0139452P.	PR	13-AUG-1999;	99US-0148684P.
PR	18-JUN-1999;	99US-0139454P.	PR	16-AUG-1999;	99US-0149368P.
PR	18-JUN-1999;	99US-0139454P.	PR	17-AUG-1999;	99US-0149175P.
PR	18-JUN-1999;	99US-0139456P.	PR	18-AUG-1999;	99US-0149426P.
PR	18-JUN-1999;	99US-0139457P.	PR	20-AUG-1999;	99US-0149722P.
PR	18-JUN-1999;	99US-0139457P.	PR	20-AUG-1999;	99US-0149723P.
PR	18-JUN-1999;	99US-0139459P.	PR	20-AUG-1999;	99US-0149902P.
PR	18-JUN-1999;	99US-0139460P.	PR	23-AUG-1999;	99US-0149902P.
PR	18-JUN-1999;	99US-0139461P.	PR	23-AUG-1999;	99US-0149930P.
PR	18-JUN-1999;	99US-0139462P.	PR	25-AUG-1999;	99US-0150566P.
PR	18-JUN-1999;	99US-0139463P.	PR	26-AUG-1999;	99US-0150884P.
PR	18-JUN-1999;	99US-0139750P.	PR	27-AUG-1999;	99US-0151065P.
PR	18-JUN-1999;	99US-0139753P.	PR	27-AUG-1999;	99US-0151066P.
PR	21-JUN-1999;	99US-0139817P.	PR	27-AUG-1999;	99US-0151080P.
PR	22-JUN-1999;	99US-0139899P.	PR	30-AUG-1999;	99US-0151303P.
PR	23-JUN-1999;	99US-0140333P.	PR	31-AUG-1999;	99US-0151438P.
PR	23-JUN-1999;	99US-0140354P.	PR	01-SEP-1999;	99US-0151930P.
PR	24-JUN-1999;	99US-0140695P.	PR	07-SEP-1999;	99US-0152363P.
PR	26-JUN-1999;	99US-0140823P.	PR	10-SEP-1999;	99US-0153070P.
PR	29-JUN-1999;	99US-0140991P.	PR	13-SEP-1999;	99US-0153758P.
PR	30-JUN-1999;	99US-0141287P.	PR	15-SEP-1999;	99US-0154018P.
PR	01-JUL-1999;	99US-0141842P.	PR	16-SEP-1999;	99US-0154039P.
PR	01-JUL-1999;	99US-0142154P.	PR	20-SEP-1999;	99US-0154779P.
PR	02-JUL-1999;	99US-0142055P.	PR	22-SEP-1999;	99US-0155139P.
PR	06-JUL-1999;	99US-0142390P.	PR	23-SEP-1999;	99US-0155486P.
PR	08-JUL-1999;	99US-0142803P.	PR	24-SEP-1999;	99US-0155659P.
PR	09-JUL-1999;	99US-0142920P.	PR	28-SEP-1999;	99US-0158458P.
PR	12-JUL-1999;	99US-0142977P.	PR	29-SEP-1999;	99US-0158596P.
PR	13-JUL-1999;	99US-0143542P.	PR	04-OCT-1999;	99US-0157117P.
PR	14-JUL-1999;	99US-0143624P.	PR	05-OCT-1999;	99US-0157753P.
PR	15-JUL-1999;	99US-0144005P.	PR	06-OCT-1999;	99US-0157865P.
PR	16-JUL-1999;	99US-0144085P.	PR	07-OCT-1999;	99US-0158029P.
PR	16-JUL-1999;	99US-0144086P.	PR	08-OCT-1999;	99US-0158232P.
PR	19-JUL-1999;	99US-0144332P.	PR	12-OCT-1999;	99US-0158369P.
PR	19-JUL-1999;	99US-0144331P.	PR	13-OCT-1999;	99US-0159293P.
PR	19-JUL-1999;	99US-0144332P.	PR	13-OCT-1999;	99US-0159294P.
PR	19-JUL-1999;	99US-0144333P.	PR	13-OCT-1999;	99US-0159295P.
PR	19-JUL-1999;	99US-0144334P.	PR	14-OCT-1999;	99US-0159329P.





```
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 23-JUL-1999; 99US-0145276P.
PR 26-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 27-JUL-1999; 99US-0145951P.
PR 28-JUL-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157533P.
PR 06-OCT-1999; 99US-0157865P.
PR 08-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158233P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159633P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.

PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161408P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 26-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 63.2%; Score 43; DB 3; Length 272;
Best Local Similarity 53.8%; Pred. NO. 9.1;
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 DVRPYRTSKYLEL 13
DB 97 DTHPHRMSRYIQL 109

RESULT 10
AAG06654
ID AAG06654 standard; protein; 281 AA.
XX
AC AAG06654;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 3505.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
hybridisation assay; genetic mapping; gene expression control; promoter;
termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EF1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-00301439.
XX
PR 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 23-APR-1999; 99US-0130891P.
PR 28-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 07-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
```

PR 14-MAY-1999; 99US-0134219P.  
PR 14-MAY-1999; 99US-0134221P.  
PR 14-MAY-1999; 99US-0134370P.  
PR 16-MAY-1999; 99US-0134768P.  
PR 19-MAY-1999; 99US-0134941P.  
PR 21-MAY-1999; 99US-0135124P.  
PR 21-MAY-1999; 99US-0135353P.  
PR 24-MAY-1999; 99US-0135629P.  
PR 25-MAY-1999; 99US-0136021P.  
PR 27-MAY-1999; 99US-0136392P.  
PR 28-MAY-1999; 99US-0137282P.  
PR 01-JUN-1999; 99US-0137222P.  
PR 03-JUN-1999; 99US-0137528P.  
PR 04-JUN-1999; 99US-0137502P.  
PR 07-JUN-1999; 99US-0137724P.  
PR 08-JUN-1999; 99US-0138094P.  
PR 10-JUN-1999; 99US-0138540P.  
PR 10-JUN-1999; 99US-0138947P.  
PR 14-JUN-1999; 99US-0139119P.  
PR 16-JUN-1999; 99US-0139452P.  
PR 16-JUN-1999; 99US-0139453P.  
PR 17-JUN-1999; 99US-0139492P.  
PR 18-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139460P.  
PR 18-JUN-1999; 99US-0139461P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 18-JUN-1999; 99US-0139763P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139899P.  
PR 23-JUN-1999; 99US-0140353P.  
PR 23-JUN-1999; 99US-0140354P.  
PR 24-JUN-1999; 99US-0140695P.  
PR 28-JUN-1999; 99US-0140823P.  
PR 30-JUN-1999; 99US-0140991P.  
PR 30-JUN-1999; 99US-0142887P.  
PR 01-JUL-1999; 99US-0142842P.  
PR 01-JUL-1999; 99US-0142154P.  
PR 02-JUL-1999; 99US-0142055P.  
PR 06-JUL-1999; 99US-0142390P.  
PR 08-JUL-1999; 99US-0142803P.  
PR 09-JUL-1999; 99US-0142920P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143542P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144085P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 19-JUL-1999; 99US-0144325P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 19-JUL-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144352P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 20-JUL-1999; 99US-0144633P.  
PR 21-JUL-1999; 99US-0144884P.  
PR 21-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145085P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145194P.  
PR 23-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 23-JUL-1999; 99US-0145224P.  
PR 26-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 27-JUL-1999; 99US-0145919P.  
PR 28-JUL-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 02-AUG-1999; 99US-0146389P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 04-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 06-AUG-1999; 99US-0147416P.  
PR 09-AUG-1999; 99US-0147493P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 10-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148319P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148565P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 16-AUG-1999; 99US-0149368P.  
PR 17-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149426P.  
PR 20-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149929P.  
PR 23-AUG-1999; 99US-0149902P.  
PR 23-AUG-1999; 99US-0149930P.  
PR 25-AUG-1999; 99US-0150566P.  
PR 26-AUG-1999; 99US-0150884P.  
PR 27-AUG-1999; 99US-0151065P.  
PR 27-AUG-1999; 99US-0151066P.  
PR 27-AUG-1999; 99US-0151080P.  
PR 30-AUG-1999; 99US-0151303P.  
PR 31-AUG-1999; 99US-0151438P.  
PR 01-SEP-1999; 99US-0151930P.  
PR 07-SEP-1999; 99US-0152163P.  
PR 10-SEP-1999; 99US-0153070P.  
PR 13-SEP-1999; 99US-0153758P.  
PR 15-SEP-1999; 99US-0154039P.  
PR 16-SEP-1999; 99US-0154018P.  
PR 20-SEP-1999; 99US-0154779P.  
PR 22-SEP-1999; 99US-0155139P.  
PR 24-SEP-1999; 99US-0155466P.  
PR 24-SEP-1999; 99US-0155659P.  
PR 28-SEP-1999; 99US-0156458P.  
PR 29-SEP-1999; 99US-0156596P.  
PR 04-OCT-1999; 99US-0157117P.  
PR 05-OCT-1999; 99US-0157753P.  
PR 06-OCT-1999; 99US-0157865P.  
PR 07-OCT-1999; 99US-0158029P.  
PR 08-OCT-1999; 99US-0158232P.  
PR 12-OCT-1999; 99US-0158369P.  
PR 13-OCT-1999; 99US-0159293P.  
PR 13-OCT-1999; 99US-0159294P.  
PR 13-OCT-1999; 99US-0159295P.  
PR 14-OCT-1999; 99US-0159329P.  
PR 14-OCT-1999; 99US-0159330P.  
PR 14-OCT-1999; 99US-0159331P.  
PR 14-OCT-1999; 99US-0159637P.  
PR 14-OCT-1999; 99US-0159638P.  
PR 18-OCT-1999; 99US-0159584P.  
PR 21-OCT-1999; 99US-0160741P.  
PR 21-OCT-1999; 99US-0160767P.  
PR 21-OCT-1999; 99US-0160768P.  
PR 21-OCT-1999; 99US-0160770P.  
PR 21-OCT-1999; 99US-0160814P.  
PR 21-OCT-1999; 99US-0160815P.  
PR 22-OCT-1999; 99US-0160980P.  
PR 22-OCT-1999; 99US-0160981P.  
PR 22-OCT-1999; 99US-0160989P.

Claim 27; SEQ ID NO 839; 458pp + Sequence Listing; English.



CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene  
CC therapy). The sequence presented is a rat protein (shown in Table 2 of  
CC the specification) which is differentially expressed during pain. Note:  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic form directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 451 AA;  
Query Match 58.8%; Score 40; DB 7; Length 451;  
Best Local Similarity 53.8%; Pred. No. 60;  
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 DVPYRTSRYLEL 13  
| | | | |  
Db 415 DYRPHKPSLYLEI 427

## RESULT 15

ABP35493  
ID ABP35493 standard; protein; 57 AA.

XX AC ABP35493;

XX 08-JUL-2002 (first entry)

XX Human ORF4466 protein, SEQ ID NO:8932.

XX Human; ORF; open reading frame; ORFX; drug screening; diagnosis;  
XX disease monitoring; cytokine; cell proliferation; cell differentiation;  
XX immune modulation; haematopoiesis regulation; tissue growth;  
XX angiogenesis; activin; inhibin; chemotactic; chemokinetic; haemostatic;  
XX thrombolytic; tumour inhibition; bodily characteristic; fertility;  
XX behaviour; cancer; proliferative disorder; neurological disorder;  
XX cardiovascular disease; immune system disorder; organ transplantation;  
XX tissue growth disorder; tissue regeneration disorder; diabetes mellitus;  
XX hypothyroidism; cholesterol ester storage disease; infection; vulnery;  
XX vasotropic; antipsoriatic; antidiabetic; cytosstatic; neurotropic;  
XX neuroprotective; antiatherosclerotic; anticoagulant; thrombolytic;  
XX cardiant; hypotensive; antithyroid; antiinflammatory; immunomodulator;  
XX dermatological; analgesic; virucide; antibacterial; fungicide.

XX Homo sapiens.

XX WO200190366-A2.

XX 29-NOV-2001.

XX 24-MAY-2001; 2001WO-US017076.

XX 24-MAY-2000; 2000US-0206690P.

XX (CURA-) CURAGEN CORP.

XX Leach MD, Shinkets RA;

XX WPI; 2002-106200/14.

XX N-PSDB; ABN79519.

XX Novel human polypeptides and polynucleotides useful for diagnosing,  
XX preventing and treating cardiovascular disease, neurodegenerative,  
XX hyperproliferative disorders and disorders related to organ  
XX transplantation.

XX Claim 10; Page 2475; 2508pp; English.

XX Sequences ABP31028-ABP35561 represent 4534 novel human proteins  
XX designated ORF (open reading frame) 1-4534, and sequences ABN75054-  
XX ABN79587 represent cDNAs encoding them. The invention also encompasses  
XX polypeptides at least 80% identical to the ORF1-ORF4534 (collectively  
XX referred to as ORFX) proteins, polynucleotides at least 85% identical to  
XX the ORFX nucleic acid sequences, vectors and host cells comprising ORFX  
XX polynucleotides, the recombinant production of ORFX proteins, antibodies

CC specific for ORFX proteins, methods of detecting ORFX polynucleotides and  
CC polypeptides, methods of screening for modulators of ORFX expression or  
CC activity, and methods of screening individuals for a predisposition to an  
CC ORFX-associated disorder. The ORFX proteins of the invention have a wide  
CC range of biological activities, such as cytokine, cell proliferation,  
CC cell differentiation, immune modulation, haematopoiesis regulation,  
CC tissue growth, angiogenesis, activin or inhibin activity, chemotactic/  
CC chemokinetic activity, haemostatic activity, thrombolytic activity,  
CC receptor/ligand, antiinflammatory activity, tumour inhibition activity,  
CC and antiinfective activity, and may also be involved in the determination  
CC of bodily characteristics, fertility and behaviour. ORFX proteins,  
CC nucleic acids and antibodies may be used in the treatment of cancers,  
CC other proliferative disorders such as psoriasis and benign tumours,  
CC neurological disorders such as epilepsy and Alzheimer's disease,  
CC cardiovascular diseases, immune system disorders, disorders related to  
CC organ transplantation, disorders of tissue growth and regeneration,  
CC diseases such as diabetes mellitus, hypothyroidism, and cholesterol ester  
CC storage disease, and infectious diseases caused by viral, bacterial,  
CC fungal and other pathogens. ORFX nucleic acids may also be used as a  
CC source of primers and probes, in the detection of ORFX genomic sequences  
CC or transcripts, in the identification and cloning of homologous  
CC sequences, in genetic diagnosis, and in forensic biology. The ORFX  
CC nucleic acids may additionally be used to produce transgenic animals  
CC which may be useful for studying the function and/or activity of ORFX  
CC protein, and in drug screening. The ORFX proteins may also be used as  
CC immunogens to generate specific antibodies, which are useful in the  
CC diagnosis, treatment and monitoring of ORFX-associated diseases

XX SQ Sequence 57 AA;

Query Match 57.4%; Score 39; DB 5; Length 57;  
Best Local Similarity 80.0%; Pred. No. 9.3;  
Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 PYRTSRYLEL 13

Db 2 PYRTSNNLEL 11

Search completed: October 5, 2004, 07:59:38  
Job time : 11:8904 secs

This Page Blank (uspto)



Best Local Similarity 66.7%; Pred. No. 21;  
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 DVPRPRTSRYLE 12  
I:|||||  
Db 429 DLRPFPNSYLE 440

RESULT 3  
SNECPII  
Pitriylisin (EC 3.4.24.55) precursor [validated] - Escherichia coli (strain K-12)  
N:Alternate names: endopeptidase P1; proteinase III  
C:Species: Escherichia coli  
C>Date: 31-Mar-1993 #sequence\_revision 31-Oct-1997 #text\_change 27-Oct-2003  
C:Accession: F5064; A29093; A25765; B25532  
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co-  
A.; Rose, D.J.; Mau, B.; Shao, Y.  
Science 277, 1453-1462, 1997  
A:Title: The complete genome sequence of Escherichia coli K-12.  
A:Reference number: A64720; MUID:97426617; PMID:9278503  
A:Accession: F5064  
A>Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-962 <ELAT>  
A:Cross-references: GB:AB000365; GB:U00096; NID:G2367163; PIDN:AAC75860.1; PID:G2367164;  
A:Experimental source: strain K-12, substrain MGL655  
R:Claverie-Martin, F.; Diaz-Torres, M.R.; Kushner, S.R.  
Gene 54, 185-195, 1987  
A:Title: Analysis of the regulatory region of the protease III (ptr) gene of Escherichia  
A:Reference number: A29093; MUID:88005781; PMID:3308636  
A:Accession: A29093  
A:Molecule type: DNA  
A:Residues: 1-276 'HYHSLR', 283 'W', 285-296 <CLA>  
A:Cross-references: GB:M7095; NID:G147390; PIDN:AAA24436.1; PID:G147391  
A:Experimental source: strain K12  
A>Note: part of this sequence, including the amino end of the mature protein, was confir-  
R:Finch, P.W.; Wilson, R.E.; Brown, K.; Hickson, I.D.; Emmerson, P.T.  
Nucleic Acids Res. 14, 7695-7703, 1986  
A:Title: Complete nucleotide sequence of the Escherichia coli ptr gene encoding protease  
A:Reference number: A25765; MUID:87040734; PMID:3534791  
A:Accession: A25765  
A:Molecule type: DNA  
A:Residues: 1-962 <FIN>  
A:Cross-references: GB:X06227; NID:G42560; PIDN:CAA29576.1; PID:G42561  
R:Becker, A.B.; Roth, R.A.  
Proc. Natl. Acad. Sci. U.S.A. 89, 3835-3839, 1992  
A:Title: An unusual active site identified in a family of zinc metalloendopeptidases.  
A:Reference number: A38854; MUID:92237263; PMID:1570301  
A:Contents: annotation; active site  
C:Genetics:  
A:Gene: ptr  
A:Map position: 61  
C:Function:  
A:Description: endopeptidase degrades small peptides [validated, MUID:92237263]  
A:Pathway: protein degradation  
C:Superfamily: insulin-degrading enzyme (IDE)  
C:Keywords: hydrolase; metalloproteinase; monomer; periplasmic space; protein degradation  
F:1-23/Domain: signal sequence #status predicted <SIG>  
F:24-962/Product: pitriylisin #status experimental <MAT>  
F:88.92/Binding site: zinc (His) #status experimental  
F:91/Active site: Glu #status experimental

Query Match 60.3%; Score 41; DB 1; Length 962;  
Best Local Similarity 70.0%; Pred. No. 26;  
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 4 PYRTSRYLEI 13  
I:|||||  
Db 124 PYRTAFYLEV 133

RESULT 4  
F91088  
protease III precursor (pitriylisin) [imported] - Salmonella enterica subsp. enterica sero-  
C:Species: Salmonella enterica subsp. enterica serovar Typhi  
A>Note: this species has also been called Salmonella typhi  
C>Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 27-Oct-2003  
C:Accession: AC0865  
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,  
th, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,  
S.; Moule, S.; O'Gaora, P.  
Nature 413, 848-852, 2001  
A:Authors: Farry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;  
A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov-  
A:Reference number: AB0502; MUID:21534947; PMID:11677608  
A:Accession: AC0865

protease III [imported] - Escherichia coli (strain O157:H7, substrain RIMD 0509952)  
C:Species: Escherichia coli  
C>Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 27-Oct-2003  
C:Accession: F91088  
R:Hayaishi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.;  
Gawawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, I.; Hattori, M.; Shinagawa, H.  
DNA Res. 8, 11-22, 2001  
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genon  
A:Reference number: A99629; MUID:21156231; PMID:11258796  
A:Accession: F91088  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-962 <HAY>  
A:Cross-references: GB:BA000007; PIDN:BA37101.1; PID:gl3363150; GSPDB:GN00154  
A:Experimental source: strain O157:H7, substrain RIMD 0509952  
C:Genetics:  
A:Gene: ECs3678  
C:Superfamily: insulin-degrading enzyme (IDE)

Query Match 60.3%; Score 41; DB 2; Length 962;  
Best Local Similarity 70.0%; Pred. No. 26;  
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 4 PYRTSRYLEI 13  
I:|||||  
Db 124 PYRTAFYLEV 133

RESULT 5  
H85933  
protease III [imported] - Escherichia coli (strain O157:H7, substrain EDL933)  
C:Species: Escherichia coli  
C>Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 27-Oct-2003  
C:Accession: H85933  
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew,  
iller, L.; Grobbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,  
Nature 409, 529-533, 2001  
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
A:Reference number: A95480; MUID:21074935; PMID:11206551  
A:Accession: H85933  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-962 <STO>  
A:Cross-references: GB:AB005174; NID:G12517303; PIDN:AAG57932.1; GSPDB:GN00145; UWMP:Z41:  
A:Experimental source: strain O157:H7, substrain EDL933  
C:Genetics:  
A:Gene: ptr  
C:Superfamily: insulin-degrading enzyme (IDE)

Query Match 60.3%; Score 41; DB 2; Length 962;  
Best Local Similarity 70.0%; Pred. No. 26;  
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 4 PYRTSRYLEI 13  
I:|||||  
Db 124 PYRTAFYLEV 133

RESULT 6  
AC0865  
protease III precursor (pitriylisin) [imported] - Salmonella enterica subsp. enterica sero-  
C:Species: Salmonella enterica subsp. enterica serovar Typhi  
A>Note: this species has also been called Salmonella typhi  
C>Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 27-Oct-2003  
C:Accession: AC0865  
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,  
th, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,  
S.; Moule, S.; O'Gaora, P.  
Nature 413, 848-852, 2001  
A:Authors: Farry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;  
A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov-  
A:Reference number: AB0502; MUID:21534947; PMID:11677608  
A:Accession: AC0865



A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-962 <PAR>  
A:Cross-references: GB:AL513382; PIDN:CAD02819.1; PID:g16504073; GSPDB:GN00176  
C:Genetics:  
C:Superfamily: insulin-degrading enzyme (IDE)  
Query Match 60.3%; Score 41; DB 2; Length 962;  
Best Local Similarity 70.0%; Pred. No. 26;  
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
QY 4 DVPRYTSRYLE 13  
Db 124 DVPRYTSRYLE 133

RESULT 7  
AD3042  
oxidoreductase Atu3952 [imported] - Agrobacterium tumefaciens (strain C58, Dupont)  
C:Species: Agrobacterium tumefaciens  
C:Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 18-Nov-2002  
C:Accession: AD3042  
R:Wood, D.W.; Scubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.B.; Chen, Y.; Woo, L.  
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kuyavin, T.; Levy, R.; Li, M.; McClell  
; Karp, P.; Romero, P.; Zhang, S.  
Science 294, 2317-2323, 2001  
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,  
ster, E.W.  
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.  
A:Reference number: AB2577; MUID:21608550; PMID:11743193  
A:Accession: AD3042  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-442 <KUR>  
A:Cross-references: GB:AE008689; PIDN:AAL44754.1; PID:g17742390; GSPDB:GN00187  
A:Experimental source: strain C58 (Dupont)  
C:Genetics:  
A:Gene: Atu3952  
A:Map position: linear chromosome  
Query Match 58.8%; Score 40; DB 2; Length 442;  
Best Local Similarity 58.3%; Pred. No. 18;  
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;  
QY 1 DVPRYTSRYLE 12  
Db 421 DVPRYTSRYLE 432

RESULT 8  
H98243  
agae protein [imported] - Agrobacterium tumefaciens (strain C58, Cereon)  
C:Species: Agrobacterium tumefaciens  
C:Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 18-Nov-2002  
C:Accession: H98243  
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,  
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;  
Science 294, 2323-2328, 2001  
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum  
A:Reference number: A97359; MUID:21608551; PMID:11743194  
A:Accession: H98243  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-442 <KUR>  
A:Cross-references: GB:AE007870; PIDN:AAK89474.1; PID:g15159342; GSPDB:GN00170  
C:Genetics:  
A:Gene: AGR\_L1801  
A:Map position: linear chromosome  
Query Match 58.8%; Score 40; DB 2; Length 442;  
Best Local Similarity 58.3%; Pred. No. 18;  
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 DVPRYTSRYLE 12  
Db 421 DVPRYTSRYLE 432

RESULT 9  
T00447  
hypothetical protein T14N5.4 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 01-Feb-1999 #sequence\_revision 01-Feb-1999 #text\_change 22-Oct-1999  
C:Accession: T00447  
R:Federspiel, N.A.; Palm, C.J.; Conway, A.B.; Kurtz, D.B.; Conway, A.R.; Au, M.; Araujo,  
; Vysotskaia, V.S.; Yu, G.; Ecker, J.; Theologis, A.; Davis, R.W.  
submitted to the EMBL Data Library, September 1998  
A:Reference number: Z14152  
A:Accession: T00447  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-258 <PED>  
A:Cross-references: EMBL:AC004260; NID:g3176694; PID:g3540203; GSPDB:GN00059; ATSP:T14N5.  
A:Experimental source: cultivar Columbia  
C:Genetics:  
A:Gene: ATSP:T14N5.4  
A:Map position: 1  
A:Introns: 38/3; 106/1  
Query Match 57.4%; Score 39; DB 2; Length 258;  
Best Local Similarity 66.7%; Pred. No. 16;  
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;  
QY 4 PYRTSRYLE 12  
Db 199 PYRTSRYLE 207

RESULT 10  
A86274  
P7A19, 15 protein - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 09-Nov-2001  
C:Accession: A86274  
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;  
ansen, N.F.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.;  
C.A.; Li, J.H.; Li, X.; Lin, X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, I  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A86141; MUID:21016719; PMID:11130712  
A:Accession: A86274  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-526 <STO>  
A:Cross-references: GB:AE005172; NID:g5080782; PIDN:AA039292.1; GSPDB:GN00141  
C:Genetics:  
A:Map position: 1  
Query Match 57.4%; Score 39; DB 2; Length 526;  
Best Local Similarity 87.5%; Pred. No. 33;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 4 PYRTSRYL 11  
Db 79 PYRTSEYL 86

RESULT 11  
B82559  
30S ribosomal protein S1 XF2438 [imported] - Xylella fastidiosa (strain 9a5c)

C:Species: Xylella fastidiosa  
 C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Sep-2000  
 C:Accession: B82559  
 R:Anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequencing  
 Nature 406, 151-157, 2000  
 A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.  
 A:Reference number: A82515; PMID:20365717; PMID:10910347  
 A:Note: for a complete list of authors see reference number A59328 below  
 A:Accession: B82559  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-569 <SIM>  
 A:Cross-references: GB:AE000452; GB:AE003849; NID:G9107617; PIDN:AAF85237.1; GSPDB:GN001  
 A:Experimental source: strain 945C  
 R:Simpson, A.J.G.; Reinach, P.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A  
 Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H  
 as-Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.  
 submitted to GenBank, June 2000  
 A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franco, M.C.; Frohn  
 J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laigh  
 chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E  
 A:Authors: Martins, E.M.F.; Matsukuma, A.V.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;  
 F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A  
 Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak  
 A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir  
 M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z  
 A:Reference number: A59328  
 A:Contents: annotation  
 C:Genetics:  
 A:Gene: XF2438  
 C:Superfamily: Escherichia coli ribosomal protein S1

Query Match 57.4%; Score 39; DB 2; Length 569;  
 Best Local Similarity 66.7%; Pred. No. 35;  
 Matches 8; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 QY 1 DVPRVTSRYLEI 12  
 |||||  
 DB 149 DVPRVDRPAYLE 160  
 |||||

RESULT 12  
 T15202  
 hypothetical protein F12B6.3 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 20-Sep-1999  
 C:Accession: T15202  
 R:Pauley, A.; Maggi, L.  
 submitted to the EMBL Data Library, May 1997  
 A:Description: The sequence of C. elegans cosmid F12B6.  
 A:Reference number: 218307  
 A:Accession: T15202  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-594 <PAU>  
 A:Cross-references: EMBL:AF003138; NID:G2088708; PID:G2088711; PIDN:AAB54155.1; GSPDB:GN  
 A:Experimental source: strain Bristol N2; clone F12B6  
 C:Genetics:  
 A:Gene: CESP:F12B6.3  
 A:Map position: 1  
 A:Introns: 45/2; 146/3; 211/3; 316/1; 367/3; 407/3; 552/2

Query Match 57.4%; Score 39; DB 2; Length 594;  
 Best Local Similarity 53.8%; Pred. No. 37;  
 Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;  
 QY 1 DVPRVTSRYLEI 13  
 |||||  
 DB 289 DLSPVTSRTEV 301  
 |||||

RESULT 13  
 T24300

ribosomal protein S14 T01E8.6, mitochondrial [similarity] - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 03-May-2002  
 C:Accession: T24300  
 R:McMurray, A.  
 submitted to the EMBL Data Library, March 1995  
 A:Reference number: Z19871  
 A:Accession: T24300  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-103 <WIL>  
 A:Cross-references: EMBL:Z48809; PIDN:CAA88746.1; GSPDB:GN00020; CESP:T01E8.6  
 A:Experimental source: clone T01E8  
 C:Genetics:  
 A:Gene: CESP:T01E8.6  
 A:Map position: 2  
 A:Introns: 20/3  
 C:Superfamily: Escherichia coli ribosomal protein S14

Query Match 55.9%; Score 38; DB 2; Length 103;  
 Best Local Similarity 60.0%; Pred. No. 9.5;  
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
 QY 2 VRPRTSRYL 11  
 |||||  
 DB 76 IKPYRLSRHL 85  
 |||||

RESULT 14  
 G90705  
 probable transport ECs0615 [imported] - Escherichia coli (strain O157:H7, substrain RMD  
 C:Species: Escherichia coli  
 C:Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 18-Jul-2001  
 C:Accession: G90705  
 R:Hayashi, T.; Makino, K.; Onishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.;  
 Gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.  
 DNA Res. 8, 11-22, 2001  
 A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genom  
 A:Reference number: A99629; PMID:21156231; PMID:11258796  
 A:Accession: G90705  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-415 <HAY>  
 A:Cross-references: GB:BA000007; PIDN:BA34038.1; PID:gl3360073; GSPDB:GN00154  
 A:Experimental source: strain O157:H7, substrain RMD 050952  
 C:Genetics:  
 A:Gene: ECs0615

Query Match 55.9%; Score 38; DB 2; Length 415;  
 Best Local Similarity 58.3%; Pred. No. 39;  
 Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;  
 QY 2 VRPRTSRYLEI 13  
 |||||  
 DB 291 LKPYLTSRHOEI 302  
 |||||

RESULT 15  
 B85556  
 probable transport ybdG [imported] - Escherichia coli (strain O157:H7, substrain EDL933)  
 C:Species: Escherichia coli  
 C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001  
 C:Accession: B85556  
 R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew,  
 Miller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamouis, K.; Apodaca,  
 Nature 409, 529-533, 2001  
 A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
 A:Reference number: A85480; PMID:21074935; PMID:11206551  
 A:Accession: B85556  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-415 <STO>  
 A:Cross-references: GB:AE005174; NID:gl2513462; PIDN:AAG54910.1; GSPDB:GN00145; UWGP:Z077

A:Experimental source: strain O157:H7, substrain EDL933

C:Genetics:

A:Gene: ybdG

Query Match 55.9%; Score 38; DB 2; Length 415;

Best Local Similarity 58.3%; Pred. No. 39;

Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 2 VRPYRTSRYLEI 13

Db 291 LXPYLTSRHQEI 302

Search completed: October 5, 2004, 08:16:50

Job time : 2.10648 secs

This Page Blank (uspic)

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2004, 07:29:35 ; Search time 1.16358 Seconds  
(without alignments)  
581.749 Million cell updates/sec

Title: US-09-805-290A-10  
Perfect score: 68  
Sequence: 1 DVPRYRTSYLEI 13

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	41	60.3	962	PTRA_ECO57	Q8x6m8 escherichia
2	41	60.3	962	PTRA_ECOL6	Q8cvs2 escherichia
3	41	60.3	962	PTRA_ECOLI	P05458 escherichia
4	41	60.3	962	PTRA_SALTI	Q8z418 salmonella
5	41	60.3	962	PTRA_SALTY	Q8zmb5 salmonella
6	41	60.3	962	PTRA_SHIFL	Q83q33 shigella fl
7	39	57.4	526	FUT7_ARATH	Q9x181 arabidopsia
8	38	55.9	199	RT14_CAEEL	P49391 caenorhabdi
9	38	55.9	415	YBDG_ECOLI	P39455 escherichia
10	38	55.9	3503	DS_DROME	Q24292 drosophila
11	37	54.4	820	SVR_BIFLO	Q8g4v2 bifidobacte
12	36	52.9	99	DEFC_AEDAE	P81603 aedes aegypt
13	36	52.9	157	LUXS_BORBU	O50164 borrelia bu
14	36	52.9	300	DHAA_MYCBO	Q9xb14 mycobacteri
15	36	52.9	300	DHAA_MYCTU	O50642 mycobacteri
16	36	52.9	378	RS1_PROSP	P14128 providencia
17	36	52.9	487	SYE2_THEMA	Q9x218 thermotoga
18	36	52.9	653	MALQ_PVEKO	Q32450 pyrococcus
19	36	52.9	1405	APCI_SCHPO	Q94666 schizosacch
20	35	51.5	210	Y14F_BP74	P13321 bacterioph
21	35	51.5	371	YKQA_CAEEL	P34306 caenorhabdi
22	35	51.5	438	RHO_HELPY	Q9z189 helicobacte
23	35	51.5	438	RHO_HELPY	P56466 helicobacte
24	35	51.5	443	CBPM_HUMAN	P14384 homo sapien
25	35	51.5	444	V111_FOMPV	Q9j5b1 fowlpox vir
26	35	51.5	499	V111_MYCHO	P43055 mycoplasma
27	35	51.5	500	FUT8_ARATH	Q9x178 arabidopsia
28	35	51.5	557	RS1_ECOLI	P02349 escherichia
29	35	51.5	566	RS1_BUCBP	Q89aj3 buchnera ap
30	35	51.5	725	RRP2_INBSI	P11136 influenza b
31	35	51.5	726	RRP2_INBAC	P13873 influenza b
32	35	51.5	726	RRP2_INBAC	P13874 influenza b
33	35	51.5	726	RRP2_INBAC	O36432 influenza b

34	35	51.5	731	1	DAXX_RAT	Q8vib2 rattus norv
35	35	51.5	739	1	DAXX_MOUSE	O35613 mus musculu
36	35	51.5	740	1	DAXX_HUMAN	Q9uer7 homo sapien
37	35	51.5	960	1	CHP1_SCHPO	Q10103 schizosacch
38	35	51.5	3056	1	ATM_HUMAN	Q13315 homo sapien
39	35	51.5	3066	1	ATM_MOUSE	O62398 mus musculu
40	34.5	50.7	349	1	MORA_RHIME	Q92pb4 rhizobium m
41	34.5	50.7	737	1	AD02_RAT	Q82b02 rattus norv
42	34	50.0	59	1	RL32_BORBU	O51646 borrelia bu
43	34	50.0	96	1	Y4KP_RHISN	P55534 rhizobium s
44	34	50.0	230	1	FLPA_THEAC	Q9hjl8 thermoplasm
45	34	50.0	251	1	VIF_FIVPE	P16089 feline immu

ALIGNMENTS

RESULT 1  
PTRA\_ECO57  
ID PTRA\_ECO57 STANDARD; PRT; 962 AA.  
AC Q8x6m8;  
DT 10-OCT-2003 (Rel. 42, Created)  
DT 10-OCT-2003 (Rel. 42, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Protease III precursor (EC 3.4.24.55) (Pitrilysin) (Protease pi).  
GN PTRA OR PTR OR 24138 OR ECS3678.  
OS Escherichia coli O157:H7.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Escherichia.  
OX NCBI\_TaxID=83334;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;  
RX MEDLINE=21074935; PubMed=11205551;  
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,  
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,  
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,  
RA Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamocousis K.,  
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,  
RA Welch R.A., Blattner F.R.;  
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7";  
RL Nature 409:529-533(2001).  
SEQUENCE FROM N.A.  
RC STRAIN=O157:H7 / RIMD 0509952;  
RX MEDLINE=21156231; PubMed=11258796;  
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,  
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,  
RA Iida T., Takami H., Honda T., Sasaki K., Ogasawara N., Yasunaga T.,  
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;  
RA "Complete genome sequence of enterohaemorrhagic Escherichia coli  
O157:H7 and genomic comparison with a laboratory strain K-12";  
RL DNA Res. 8:11-22(2001).  
CC -!- FUNCTION: Endopeptidase that degrades small peptides of less than  
7 kDa, he such as glucagon and insulin (By similarity).  
CC -!- CATALYTIC ACTIVITY: Preferential cleavage of 16-Tyr-|-Leu-17 and  
25-Phe-|-Tyr-26 bonds of oxidized insulin B chain. Also acts on  
other substrates of Mw less than 7 kDa such as insulin and  
glucagon.  
CC -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).  
CC -!- SUBUNIT: Monomer (By similarity).  
CC -!- SUBCELLULAR LOCATION: Periplasmic (By similarity).  
CC -!- SIMILARITY: Belongs to peptidase family M16.

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).  
-----  
EMBL; A5005510; AAG57932.1; -.

DR EMBL; AF002563; BAB37101.1; -  
 DR PIR; F91088; F91088; Peptidase M16;  
 DR InterPro; IPR001431; Peptidase M16;  
 DR InterPro; IPR007863; Peptidase M16; 1.  
 DR Pfam; PF00675; Peptidase M16; 1.  
 DR Pfam; PF05193; Peptidase M16; 2.  
 DR PROSITE; PS00143; INSULINASE; 1.  
 KW Hydrolase; Metalloprotease; Magnesium; Periplasmic; Zinc; Signal;  
 KW Complete proteome.  
 FT SIGNAL 1 23 BY SIMILARITY.  
 FT CHAIN 24 962 PROTEASE III.  
 FT METAL 88 88 ZINC (BY SIMILARITY).  
 FT ACT\_SITE 91 91 BY SIMILARITY.  
 FT METAL 92 92 ZINC (BY SIMILARITY).  
 FT METAL 169 169 ZINC (BY SIMILARITY).  
 SQ SEQUENCE 962 AA; 107808 MW; 30785644CBF0E452 CRC64;  
 Query Match 60.3%; Score 41; DB 1; Length 962;  
 Best Local Similarity 70.0%; Pred. No. 11;  
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 QY 4 PYRTSRYLEI 13  
 DB 124 PYRTAFYLEV 133  
 RESULT 2  
 ID PTR\_A\_ECOLI STANDARD; PRT; 962 AA.  
 AC Q8CVS2;  
 DT 10-OCT-2003 (Rel. 42, Created)  
 DT 10-OCT-2003 (Rel. 42, Last sequence update)  
 DE Protease III precursor (EC 3.4.24.55) (Pitrilysin) (Protease pi).  
 GN PTR\_A OR PTR OR C3415.  
 OS Escherichia coli O6.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Escherichia.  
 OX NCBI\_TaxID=217592;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C6:H1 / CFT073 / ATCC 700928;  
 RX MEDLINE=22388234; PubMed=12471157;  
 RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roséch P.,  
 RA Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,  
 RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,  
 RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;  
 RT "Extensive mosaic structure revealed by the complete genome sequence  
 of uropathogenic Escherichia coli.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024 (2002).  
 CC -!- FUNCTION: Endopeptidase that degrades small peptides of less than  
 7 kDa, such as glucagon and insulin (By similarity).  
 CC -!- CATALYTIC ACTIVITY: Preferential cleavage of 16-Tyr-|-Leu-17 and  
 25-Phe-|-Tyr-26 bonds of oxidized insulin B chain. Also acts on  
 other substrates of Mw less than 7 kDa such as insulin and  
 glucagon.  
 CC -!- COPACATOR: Binds 1 zinc ion per subunit (By similarity).  
 CC -!- SUBUNIT: Monomer (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Periplasmic (By similarity).  
 CC -!- SIMILARITY: Belongs to peptidase family M16.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 or send an email to [license@sb-sib.ch](mailto:license@sb-sib.ch)).  
 CC -----  
 DR EMBL; AE016765; AAN81860.1; -  
 DR InterPro; IPR001431; Peptidase M16;  
 DR InterPro; IPR007863; Peptidase M16; 1.  
 DR Pfam; PF00675; Peptidase M16; 1.

DR Pfam; PF05193; Peptidase M16; 2.  
 DR PROSITE; PS00143; INSULINASE; 1.  
 KW Hydrolase; Metalloprotease; Magnesium; Periplasmic; Zinc; Signal;  
 KW Complete proteome.  
 FT SIGNAL 1 23 BY SIMILARITY.  
 FT CHAIN 24 962 PROTEASE III.  
 FT METAL 88 88 ZINC (BY SIMILARITY).  
 FT ACT\_SITE 91 91 BY SIMILARITY.  
 FT METAL 92 92 ZINC (BY SIMILARITY).  
 FT METAL 169 169 ZINC (BY SIMILARITY).  
 SQ SEQUENCE 962 AA; 107891 MW; DEDD2CA2A9AADF8D CRC64;  
 Query Match 60.3%; Score 41; DB 1; Length 962;  
 Best Local Similarity 70.0%; Pred. No. 11;  
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 QY 4 PYRTSRYLEI 13  
 DB 124 PYRTAFYLEV 133  
 RESULT 3  
 ID PTR\_A\_ECOLI STANDARD; PRT; 962 AA.  
 AC P05458; P78106;  
 DT 01-NOV-1988 (Rel. 09, Created)  
 DT 01-NOV-1988 (Rel. 09, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Protease III precursor (EC 3.4.24.55) (Pitrilysin) (Protease pi).  
 GN PTR\_A OR PTR OR B2821.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Escherichia.  
 OX NCBI\_TaxID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=87040734; PubMed=3534791;  
 RA Finch P.W., Wilson R.E., Brown K., Hickson I.D., Emmerson P.T.;  
 RT "Complete nucleotide sequence of the Escherichia coli ptr gene  
 encoding protease III.";  
 RL Nucleic Acids Res. 14:7695-7703 (1986).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=KL2 / MGI655;  
 RX MEDLINE=97426617; PubMed=9278503;  
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 RA Mau B., Shao Y.;  
 RT "The complete genome sequence of Escherichia coli K-12.";  
 RL Science 277:1453-1474 (1997).  
 RN [3]  
 RP SEQUENCE OF 853-962 FROM N.A.  
 RX MEDLINE=87066729; PubMed=3537960;  
 RA Finch P.W., Storey A., Chapman K.B., Brown K., Hickson I.D.,  
 RA Emmerson P.T.;  
 RT "Complete nucleotide sequence of the Escherichia coli recB gene.";  
 RL Nucleic Acids Res. 14:8573-8582 (1986).  
 RN [4]  
 RP SEQUENCE OF 1-296 FROM N.A.  
 RC STRAIN=KL2;  
 RX MEDLINE=88005781; PubMed=3308636;  
 RA Claverie-Martin F., Diaz-Torres M.R., Kushner S.R.;  
 RT "Analysis of the regulatory region of the protease III (ptr) gene of  
 Escherichia coli K-12.";  
 RL Gene 54:185-195 (1987).  
 RN [5]  
 RP MUTAGENESIS, AND ACTIVE SITE.  
 RX MEDLINE=92237263; PubMed=1570301;  
 RA Becker A.B., Roth R.A.;  
 RT "An unusual active site identified in a family of zinc  
 metalloendopeptidases.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 89:3835-3839 (1992).

[6]  
RP MUTAGENESIS, AND ACTIVE SITE.  
RX MEDLINE=93277449; PubMed=8099278;  
RA Becker A.B., Roth R.A.,  
RT Identification of Glutamate-169 as the third zinc-binding residue in  
RT proteinase III, a member of the family of insulin-degrading  
RT enzymes.;  
RL Biochem. J. 292:137-142(1993).  
CC -!- FUNCTION: Endopeptidase that degrades small peptides of less than  
CC 7 kDa, such as Glucagon and insulin.  
CC -!- CATALYTIC ACTIVITY: Preferential cleavage of 16-Tyr-[Leu-17 and  
CC 25-Phe]-Tyr-26 bonds of oxidized insulin B chain. Also acts on  
CC other substrates of Mw less than 7 kDa such as insulin and  
CC glucagon.  
CC -!- COFACTOR: Binds 1 zinc ion per subunit.  
CC -!- SUBUNIT: Monomer.  
CC -!- SUBCELLULAR LOCATION: Periplasmic.  
CC -!- SIMILARITY: Belongs to peptidase family M16.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; X04581; CAA28249.1; -  
CC EMBL; U29581; AAB40468.1; -  
CC EMBL; A8000365; AAC75860.1; -  
CC EMBL; X06227; CAA29576.1; -  
CC EMBL; M17095; AAB24436.1; -  
CC F1R; F65064; SNECF1.  
CC ECO2DBASE; G095.0; 6TH EDITION.  
CC EcoGene; EGI0786; ptrA.  
CC InterPro; IPR001431; Peptidase\_M16.  
CC InterPro; IPR007863; Peptidase\_M16\_C.  
CC Pfam; PF00675; Peptidase\_M16; 1; 2.  
CC Pfam; PF05193; Peptidase\_M16\_C; 2.  
CC PROSITE; PS00143; INSULINASE; 1.  
CC Hydrolase; Metalloprotease; Magnesium; Zinc; Signal;  
KW Complete proteome.  
KW SIGNAL 1 23  
FT CHAIN 24 962 PROTEASE III.  
FT METAL 88 88 ZINC.  
FT ACT\_SITE 91 91  
FT METAL 92 92 ZINC.  
FT METAL 169 169 ZINC.  
FT MUTAGEN 88 88 H-R: LOSS OF ACTIVITY AND OF ZN-BINDING.  
FT MUTAGEN 91 91 E-Q: LOSS OF ACTIVITY.  
FT MUTAGEN 92 92 H-R: LOSS OF ACTIVITY AND OF ZN-BINDING.  
FT MUTAGEN 162 162 E-Q: 20% LOSS OF ACTIVITY.  
FT MUTAGEN 169 169 E-Q: LOSS OF ACTIVITY AND OF ZN-BINDING.  
FT MUTAGEN 204 204 E-Q: NO LOSS OF ACTIVITY.  
FT CONFLICT 277 284 ITHVPA -> HYSLRPW (IN REF. 4).  
SQ SEQUENCE 962 AA; 107708 MW; 0558C68C2F1A0540 CRC64;  
  
Query Match 60.3%; Score 41; DB 1; Length 962;  
Best Local Similarity 70.0%; Pred.No. 11;  
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
  
QY 4 PYRTSRYLEI 13  
DB 124 PYRTAFVLEV 133  
|||||  
  
RESULT 4  
ID\_PTRA\_SALTI STANDARD; PRT; 962 AA.  
AC Q82418;  
DT 10-OCT-2003 (Rel. 42, Created)  
DT 10-OCT-2003 (Rel. 42, Last sequence update)

10-OCT-2003 (Rel. 42, Last annotation update)  
Protease III precursor (EC 3.4.24.55) (Pitrilysin) (Protease pi).  
PTRA\_OR PTR OR STV3133 OR T2903.  
OS Salmonella typhi.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Salmonella.  
OX NCBI\_TaxID=601;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CT18;  
RX MEDLINE=21534947; PubMed=11677608;  
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,  
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,  
RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,  
RA Cronin A., Davis P., Davies R.M., Dowd I., White N., Farrar J.,  
RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,  
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,  
RA Quail M.A., Rutherford K., Simmonds M., Skelton J., Stevens K.,  
RA Whitehead S., Barrall B.G.;  
RT "Complete genome sequence of a multiple drug resistant Salmonella  
RT enterica serovar Typhi CT18";  
RN Nature 413:848-852(2001).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Ty2 / ATCC 700931;  
RX MEDLINE=22531367; PubMed=12644504;  
RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,  
RA Burland W., Kodyanni V., Schwartz D.C., Blattner F.R.;  
RT "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2  
RT and CT18";  
RL J. Bacteriol. 185:2330-2337(2003).  
CC -!- FUNCTION: Endopeptidase that degrades small peptides of less than  
CC 7 kDa, such as glucagon and insulin (By similarity).  
CC -!- CATALYTIC ACTIVITY: Preferential cleavage of 16-Tyr-[Leu-17 and  
CC 25-Phe]-Tyr-26 bonds of oxidized insulin B chain. Also acts on  
CC other substrates of Mw less than 7 kDa such as insulin and  
CC glucagon.  
CC -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).  
CC -!- SUBUNIT: Monomer (By similarity).  
CC -!- SUBCELLULAR LOCATION: Periplasmic (By similarity).  
CC -!- SIMILARITY: Belongs to peptidase family M16.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; AL627277; CAD02819.1; -  
CC EMBL; AB016843; AAO70457.1; -  
CC MEROPS; M16.001; -  
CC InterPro; IPR001431; Peptidase\_M16.  
CC InterPro; IPR007863; Peptidase\_M16\_C.  
CC Pfam; PF00675; Peptidase\_M16; 1; 2.  
CC Pfam; PF05193; Peptidase\_M16\_C; 2.  
CC PROSITE; PS00143; INSULINASE; FALSE\_NEG.  
CC Hydrolase; Metalloprotease; Magnesium; Zinc; Signal;  
KW Complete proteome.  
KW SIGNAL 1 23 BY SIMILARITY.  
FT CHAIN 24 962 PROTEASE III.  
FT METAL 88 88 ZINC (BY SIMILARITY).  
FT ACT\_SITE 91 91 BY SIMILARITY.  
FT METAL 92 92 ZINC (BY SIMILARITY).  
FT METAL 169 169 ZINC (BY SIMILARITY).  
SQ SEQUENCE 962 AA; 107524 MW; 5745C2C77F7F5832 CRC64;  
  
Query Match 60.3%; Score 41; DB 1; Length 962;  
Best Local Similarity 70.0%; Pred.No. 11;  
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
  
QY 4 PYRTSRYLEI 13

```

Db      124 PYRTAFYLEV 133
|||||: |||:
RESULT 5
PTRA_SALTY STANDARD; PRT; 962 AA.
AC Q8ZMB5;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DE Protease III precursor (EC 3.4.24.55) (Pitrilysin) (Protease pil).
GN PTRA OR PTR OR STW2995.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Forwellik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2."
RL Nature 413:852-856 (2001).
CC -!- FUNCTION: Endopeptidase that degrades small peptides of less than
CC 7 kDa, such as glucagon and insulin (By similarity).
CC -!- CATALYTIC ACTIVITY: Preferential cleavage of 16-Tyr-|-Leu-17 and
CC 25-Phe-|-Tyr-26 bonds of oxidized insulin B chain. Also acts on
CC other substrates of Mw less than 7 kDa such as insulin and
CC glucagon.
CC -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC -!- SUBUNIT: Monomer (By similarity).
CC -!- SUBCELLULAR LOCATION: Periplasmic (By similarity).
CC -!- SIMILARITY: Belongs to peptidase family M16.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE008837; AAL21871.1; -
CC STYGene; SG22222; PTRA.
CC InterPro; IPR001431; Peptidase M16.
CC InterPro; IPR007863; Peptidase M16_C.
CC Pfam; PF00675; Peptidase M16; 1.
CC Pfam; PF05193; Peptidase M16_C; 2.
CC PROSITE; PS00143; INSULINASE; 1.
CC Complete proteome.
CC
CC SIGNAL 1 23 BY SIMILARITY.
CC CHAIN 24 962 PROTEASE III.
CC FT METAL 88 88 ZINC (BY SIMILARITY).
CC FT ACT SITE 91 91 BY SIMILARITY.
CC FT METAL 92 92 ZINC (BY SIMILARITY).
CC FT METAL 169 169 ZINC (BY SIMILARITY).
CC SEQUENCE 962 AA; 107486 MW; 322AD6287B873952 CRC64;
KW Hydrolase; Metalloprotease; Magnesium; Periplasmic; Zinc; Signal;
Complete proteome.
FT SIGNAL 1 23 BY SIMILARITY.
FT CHAIN 24 962 PROTEASE III.
FT FT METAL 88 88 ZINC (BY SIMILARITY).
FT FT ACT SITE 91 91 BY SIMILARITY.
FT FT METAL 92 92 ZINC (BY SIMILARITY).
FT FT METAL 169 169 ZINC (BY SIMILARITY).
FT SEQUENCE 962 AA; 107770 MW; 662041ASDCAF254 CRC64;
Query Match 60.3%; Score 41; DB 1; Length 962;
Best Local Similarity 70.0%; Pred. No. 11;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Oy 4 PYRTSRYLEI 13
|||||: |||:
Db      124 PYRTAFYLEV 133

```

```

RESULT 6
PTRA_SHIFL STANDARD; PRT; 962 AA.
AC Q83QC3;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DE Protease III precursor (EC 3.4.24.55) (Pitrilysin) (Protease pil).
GN PTRA OR PTR OR SF2832 OR S3029.
OS Shigella flexneri.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Shigella.
OX NCBI_TaxID=623;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=301 / Serotype 2a;
RX MEDLINE=22272406; PubMed=12384590;
RA Jin O., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,
RA Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,
RA Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,
RA Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,
RA Yu J.;
RT "Genome sequence of Shigella flexneri 2a: insights into pathogenicity
RT through comparison with genomes of Escherichia coli K12 and O157."
RL Nucleic Acids Res. 30:4432-4441(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=2457T / ATCC 700930 / Serotype 2a;
RX MEDLINE=22590274; PubMed=12704152;
RA Wei J., Goldberg M.B., Burland V., Venkatesan M.M., Deng W.,
RA Fournier G., Mayhew G.F., Plunkett G. III, Rose D.J., Darling A.,
RA Mau B., Perna N.T., Payne S.M., Runyen-Janecky L.J., Zhou S.,
RA Schwartz D.C., Blattner F.R.;
RT "Complete genome sequence and comparative genomics of Shigella
RT flexneri serotype 2a strain 2457T."
RL Infect. Immun. 71:2775-2786 (2003).
CC -!- FUNCTION: Endopeptidase that degrades small peptides of less than
CC 7 kDa, such as glucagon and insulin (By similarity).
CC -!- CATALYTIC ACTIVITY: Preferential cleavage of 16-Tyr-|-Leu-17 and
CC 25-Phe-|-Tyr-26 bonds of oxidized insulin B chain. Also acts on
CC other substrates of Mw less than 7 kDa such as insulin and
CC glucagon.
CC -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC -!- SUBUNIT: Monomer (By similarity).
CC -!- SUBCELLULAR LOCATION: Periplasmic (By similarity).
CC -!- SIMILARITY: Belongs to peptidase family M16.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE015297; AAN44319.1; -
CC EMBL; AE016987; AAP18144.1; -
CC InterPro; IPR001431; Peptidase M16.
CC InterPro; IPR007863; Peptidase M16_C.
CC Pfam; PF00675; Peptidase M16; 1.
CC Pfam; PF05193; Peptidase M16_C; 2.
CC PROSITE; PS00143; INSULINASE; 1.
CC Complete proteome.
CC
CC SIGNAL 1 23 BY SIMILARITY.
CC CHAIN 24 962 PROTEASE III.
CC FT METAL 88 88 ZINC (BY SIMILARITY).
CC FT ACT SITE 91 91 BY SIMILARITY.
CC FT METAL 92 92 ZINC (BY SIMILARITY).
CC FT METAL 169 169 ZINC (BY SIMILARITY).
CC SEQUENCE 962 AA; 107770 MW; 662041ASDCAF254 CRC64;
KW Hydrolase; Metalloprotease; Magnesium; Periplasmic; Zinc; Signal;
Complete proteome.
FT SIGNAL 1 23 BY SIMILARITY.
FT CHAIN 24 962 PROTEASE III.
FT FT METAL 88 88 ZINC (BY SIMILARITY).
FT FT ACT SITE 91 91 BY SIMILARITY.
FT FT METAL 92 92 ZINC (BY SIMILARITY).
FT FT METAL 169 169 ZINC (BY SIMILARITY).
FT SEQUENCE 962 AA; 107770 MW; 662041ASDCAF254 CRC64;

```



Query Match 60.3%; Score 41; DB 1; Length 962;  
Best Local Similarity 70.0%; Pred. No. 11;  
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 4 PYRTSYL 13  
DB 124 PYRTAPYLEV 133

RESULT 7  
FUT7 ARATH STANDARD; PRT; 526 AA.  
AC Q9X181; Q9L1M2;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DE Probable fucosyltransferase 7 (EC 2.4.1.-) (AtFUT7).  
GN FUT7 OR At1G14070 OR F7A19.15 OR F6A14.19 OR F6A14.28.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; rosids;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
SEQUENCE FROM N.A.  
RC STRAIN=cv. Columbia;  
RX MEDLINE=21016719; PubMed=11130712;  
RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,  
RA White O., Alonso J., Altafi H., Araujo R., Bowman C.L., Brooks S.Y.,  
RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,  
RA Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,  
RA Dunn P., Egu P., Feldblum T.V., Feng J.-D., Fong B., Fujii C.Y.,  
RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,  
RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,  
RA Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,  
RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,  
RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziani A.,  
RA Millesich J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,  
RA Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,  
RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,  
RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,  
RA Uterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,  
RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;  
RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis  
thaliana";  
RL Nature 408:816-820 (2000).  
RN [2]  
IDENTIFICATION AS PUTATIVE FUCOSYLTRANSFERASE, AND TISSUE SPECIFICITY.  
RX MEDLINE=21608393; PubMed=11743104;  
RA Sarria B., Wagner T.A., O'Neill M.A., Faik A., Walkerson C.G.,  
RA Keegstra K., Raikhel N.V.;  
RT "Characterization of a family of Arabidopsis genes related to  
xyloglucan fucosyltransferase";  
RL Plant Physiol. 127:1595-1606 (2001).  
CC -!- FUNCTION: May be involved in cell wall biosynthesis. May  
act as a fucosyltransferase.  
CC -!- PATHWAY: Glycosylation.  
CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Membrane-bound  
form in trans cisternae of Golgi (by similarity).  
CC -!- TISSUE SPECIFICITY: Expressed in roots, leaves, stems and  
seedlings.  
CC -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 37.  
CC -!- CAUTION: Ref.1 (AAF79408) sequence differs from that shown due to  
erroneous gene model prediction.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----

EMBL; AC007576; AAD39292.1; -;  
EMBL; AC068197; AAF79408.1; ALT\_SEQ.  
PIR; A86274; A86274.  
DR InterPro: IPR004938; XG FTase.  
DR Pfam: PF03254; XG FTase; 1.  
KW Transferase; Glycosyltransferase; Transmembrane; Glycoprotein;  
FT SIGNAL-ANCHOR; Golgi stack; Cell wall.  
FT DOMAIN 1 4 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 5 25 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)  
(POTENTIAL).  
FT DOMAIN 26 526 LUMENAL, CATALYTIC (POTENTIAL).  
FT CARBOHYD 211 211 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 215 215 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 363 363 N-LINKED (GLCNAC. .) (POTENTIAL).  
SQ SEQUENCE 526 AA; 60537 MW; 8F3B447551025FCD CRC64;  
Query Match 57.4%; Score 39; DB 1; Length 526;  
Best Local Similarity 87.5%; Pred. No. 13;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 PYRTSYL 11  
DB 79 PYRTSEYL 86

RESULT 8  
RT14 CAEL STANDARD; PRT; 199 AA.  
ID RT14 CAEL  
AC P49391;  
DT 01-FEB-1996 (Rel. 33, Created)  
DT 10-OCT-2003 (Rel. 42, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Putative mitochondrial 40S ribosomal protein S14.  
DT 1088.6.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Pelodierinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RA McMurray A.A.;  
RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.  
RN [2]  
REVIEWS.  
RA Durbin R.;  
RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.  
CC -!- SUBCELLULAR LOCATION: Mitochondrial (by similarity).  
CC -!- SIMILARITY: Belongs to the S14P family of ribosomal proteins.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; Z48809; CAA88746.2; -;  
PIR; T24300; T24300.  
DR WormPep; T01E8.6; CE32923.  
DR InterPro; IPR001209; Ribosomal\_S14.  
DR Pfam; PF00253; Ribosomal\_S14; 1.  
DR PROSITE; PS00527; RIBOSOMAL\_S14; FALSE\_NEG.  
KW Ribosomal protein; Mitochondrion.  
SQ SEQUENCE 199 AA; 23414 MW; 2CC0021A618BCFF0 CRC64;  
Query Match 55.9%; Score 38; DB 1; Length 199;  
Best Local Similarity 60.0%; Pred. No. 7;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 VPRPRTSYL 11  
DB 79 PYRTSEYL 86

Db 172 IKPRLSRHL 181

RESULT 9

YBDG\_ECOLI STANDARD; PRT; 415 AA.

AC P39455; P77602;

DT 01-FEB-1995 (Rel. 31, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Hypothetical protein ybdg.

GN YBDG OR B0577 OR SF0483 OR S0492.

OS Escherichia coli, and

OS Shigella flexneri.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

OC Enterobacteriaceae; Escherichia.

OX NCBI\_TaxID=562, 623;

RN [1]

RP SEQUENCE FROM N.A.

RC SPECIES=E.coli; STRAIN=K12 / MG1655;

RX MEDLINE=97426617; PubMed=9278503;

RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,

RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,

RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,

RA Mau B., Shao Y.;

RT "The complete genome sequence of Escherichia coli K-12.";

RL Science 277:1453-1474(1997).

RN [2]

RP SEQUENCE FROM N.A.

RC SPECIES=E.coli; STRAIN=K12 / MG1655;

RX MEDLINE=97061202; PubMed=8905232;

RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,

RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,

RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,

RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,

RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,

RA Yano M., Horiuchi T.;

RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome

RT corresponding to the 12.7-28.0 min region on the linkage map.";

RL DNA Res. 3:137-155(1996).

RN [4]

RP SEQUENCE OF 1-77 FROM N.A.

RC SPECIES=E.coli; STRAIN=B;

RX MEDLINE=95113294; PubMed=7813889;

RA Michael N.P., Brenn J.K., Anlezark G.M., Minton N.P.;

RT "Physical characterisation of the Escherichia coli B gene encoding

RT nitroreductase and its over-expression in Escherichia coli K12.";

RL FEMS Microbiol. Lett. 124:195-202(1994).

RN [5]

RP SEQUENCE OF 1-12 FROM N.A.

RC SPECIES=E.coli; STRAIN=K12 / C600;

RA Zemo S., Koike H., Tanokura M., Saigo K.;

RL Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.

RN [6]

RP SEQUENCE FROM N.A.

RC SPECIES=S.flexneri; STRAIN=301 / Serotype 2a;

RX MEDLINE=22272406; PubMed=12384590;

RA Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,

RA Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,

RA Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,

RA Xu J.;

RT "Genome sequence of Shigella flexneri 2a: insights into pathogenicity

RT through comparison with genomes of Escherichia coli K12 and O157.";

RL Nucleic Acids Res. 30:4432-4441(2002).

RN [7]

RP SEQUENCE FROM N.A.

RC SPECIES=S.flexneri; STRAIN=2457T / ATCC 700930 / Serotype 2a;

RX MEDLINE=22590274; PubMed=12704152;

RA Wei J., Goldberg M.B., Burland V., Venkatesan M.M., Deng W.,

RA Fournier G., Mayhew G.F., Plunkett G. III, Rose D.J., Darling A.,

RA Mau B., Perna N.T., Payne S.M., Runyen-Janecky L.J., Zhou S.,

RA Schwartz D.C., Blattner F.R.;

RT "Complete genome sequence and comparative genomics of Shigella

RT flexneri serotype 2a strain 2457T.";

RL Infect. Immun. 71:2775-2786(2003).

CC [-] SUBCELLULAR LOCATION: Integral membrane protein (Potential).

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC

EMBL; AE000163; AAC73678.1; -

EMBL; U82598; AAB40775.1; -

DR EMBL; D90700; AAC35217.1; -

DR EMBL; U07860; AAC43265.1; -

DR EMBL; D25414; -; NOT ANNOTATED CDS.

DR EMBL; AE015080; AAN42134.1; ALT\_INIT.

DR EMBL; AE016979; AAP16008.1; -

DR PIR; G64790; G64790.

DR EcGene; EGI2620; ybdg.

DR InterPro; IPR006685; MSion\_channel.

DR Pfam; PF00924; MS\_channel; 1.

KW Hypothetical protein; Transmembrane; Complete proteome.

FT TRANSMEM 25 45 POTENTIAL.

FT TRANSMEM 68 88 POTENTIAL.

FT TRANSMEM 105 125 POTENTIAL.

FT TRANSMEM 149 169 POTENTIAL.

FT TRANSMEM 174 194 POTENTIAL.

FT TRANSMEM 370 390 POTENTIAL.

SQ SEQUENCE 415 AA; 46602 MW; C9A8A9A41CDC8F28 CRC64;

Query Match 55.9%; Score 38; DB 1; Length 415;

Best Local Similarity 58.3%; Pred No. 15;

Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 2 VRPYTSRYLEI 13

DB 291 LKPYLTSRQEI 302

RESULT 10

DS\_DROME STANDARD; PRT; 3503 AA.

ID DS\_DROME Q9VPS4;

AC Q24292; Q9VPS4;

DT 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Dachsous protein precursor (Adherin).

DE DS OR CGI7941.

GN Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI\_TaxID=7227;

RN [1]

RP SEQUENCE FROM N.A., FUNCTION, TISSUE SPECIFICITY, AND DEVELOPMENTAL

RP STAGE.

RP TISSUE=Embryo;

RX MEDLINE=95324813; PubMed=7601355;

RA Clark H.F., Brentrup D., Schneitz K., Bieber A., Goodman C., Noll M.;

RT "Dachsous encodes a member of the cadherin superfamily that controls

RT imaginal disc morphogenesis in Drosophila.";

RL Genome Dev. 9:1530-1542(1995).

RA Noll M.;  
 RL Submitted (MAY-1999) to the EMBL/GenBank/DBSJ databases.  
 RN [3]  
 RC SEQUENCE FROM N.A.  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.F., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Randell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazer D.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfankoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Banos P.V., Berkman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Borchan M.R., Buck J., Brockstein P., Brotier P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Folsler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of *Drosophila melanogaster*."  
 RL Science 287:2185-2195(2000).  
 RN [4]  
 RP REVISIONS.  
 RX MEDLINE=22426069; PubMed=12537572;  
 RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,  
 RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochuk S.E.,  
 RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,  
 RA Bettencourt B.R., Celisner S.E., de Grey A.D.N.J., Drysdale R.A.,  
 RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,  
 RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,  
 RA Lewis S.E.;  
 RT "Annotation of the *Drosophila melanogaster* euchromatic genome: a  
 systematic review".  
 RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).  
 CC -!- FUNCTION: Involved in morphogenesis. May also be involved in cell  
 CC adhesion.  
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein (Potential).  
 CC -!- TISSUE SPECIFICITY: Expressed in embryonic ectoderm. In larvae,  
 CC expression is restricted to imaginal disks and brain.  
 CC -!- DEVELOPMENTAL STAGE: Expressed throughout embryogenesis where it  
 CC is first detected during gastrulation. Also expressed in larvae  
 CC and adults.  
 CC -!- SIMILARITY: Contains 27 cadherin domains.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; L08811; AAA79329.2; -  
 DR EMBL; AB003588; AAF51468.3; ALT\_INIT.  
 DR FLYBASE; F51116; INCUJ.  
 DR FLYBASE; FBGN000497; ds.  
 DR GO; GO:0005887; C:integral to plasma membrane; ISS.  
 DR GO; GO:0008014; P:calcium-dependent cell adhesion molecule ac.; ISS.  
 DR GO; GO:0015339; P:calcium-dependent cell-cell adhesion; ISS.  
 DR GO; GO:0008283; P:cell proliferation; IMP.  
 DR GO; GO:0000904; P:cellular morphogenesis during differentiation; IMP.  
 DR GO; GO:0045317; P:equator specification; IMP.  
 DR GO; GO:0045198; P:establishment of epithelial cell polarity; IMP.  
 DR GO; GO:0018149; P:protein-protein cross-linking; IPI.  
 DR InterPro; IPR002126; Cadherin.  
 DR InterPro; IPR002233; Cadherin\_C\_term.  
 DR Pfam; PF00028; cadherin; 26  
 DR Pfam; PF01049; Cadherin\_C\_term; 1.  
 DR PRINTS; PR00205; CADHERIN.  
 DR SMART; SM00112; CA; 25.  
 DR PROSITE; PS00232; CADHERIN 1; 20.  
 DR PROSITE; PS0268; CADHERIN 2; 27.  
 KW Cell adhesion; glycoprotein; Transmembrane; Calcium; Calcium-binding;  
 KW Repeat; Signal; Developmental protein.  
 FT SIGNAL 1 20  
 FT CHAIN 21 3503  
 FT DOMAIN 21 3045 DACHSUS PROTEIN.  
 FT TRANSMEM 3046 3066 EXTRACELLULAR (POTENTIAL).  
 FT DOMAIN 3067 3503 POTENTIAL.  
 FT DOMAIN 22 121 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 122 233 CADHERIN 1.  
 FT DOMAIN 234 340 CADHERIN 2.  
 FT DOMAIN 345 451 CADHERIN 3.  
 FT DOMAIN 452 558 CADHERIN 4.  
 FT DOMAIN 559 662 CADHERIN 5.  
 FT DOMAIN 663 774 CADHERIN 6.  
 FT DOMAIN 775 878 CADHERIN 7.  
 FT DOMAIN 879 983 CADHERIN 8.  
 FT DOMAIN 984 1100 CADHERIN 9.  
 FT DOMAIN 1101 1203 CADHERIN 10.  
 FT DOMAIN 1205 1312 CADHERIN 11.  
 FT DOMAIN 1313 1432 CADHERIN 12.  
 FT DOMAIN 1433 1549 CADHERIN 13.  
 FT DOMAIN 1556 1666 CADHERIN 14.  
 FT DOMAIN 1667 1794 CADHERIN 15.  
 FT DOMAIN 1796 1899 CADHERIN 16.  
 FT DOMAIN 1900 2004 CADHERIN 17.  
 FT DOMAIN 2005 2111 CADHERIN 18.  
 FT DOMAIN 2114 2269 CADHERIN 19.  
 FT DOMAIN 2270 2375 CADHERIN 20.  
 FT DOMAIN 2375 2479 CADHERIN 21.  
 FT DOMAIN 2489 2595 CADHERIN 22.  
 FT DOMAIN 2596 2699 CADHERIN 23.  
 FT DOMAIN 2701 2809 CADHERIN 24.  
 FT DOMAIN 2810 2916 CADHERIN 25.  
 FT DOMAIN 2919 3028 CADHERIN 26.  
 FT DOMAIN 3028 3038 CADHERIN 27.  
 FT CONFLICT 1070 1070 V -> I (IN REF. 1).  
 FT CONFLICT 1490 1490 R -> S (IN REF. 1).  
 FT CONFLICT 1636 1636 S -> S (IN REF. 1).  
 FT CONFLICT 1692 1692 S -> P (IN REF. 1).  
 FT CONFLICT 1804 1804 V -> I (IN REF. 1).  
 FT CONFLICT 2029 2029 L -> I (IN REF. 1).  
 FT CONFLICT 2210 2210 P -> A (IN REF. 1).  
 FT CONFLICT 2289 2289 A -> S (IN REF. 1).  
 FT CONFLICT 2536 2536 S -> T (IN REF. 1).  
 FT CONFLICT 2862 2862 R -> Q (IN REF. 1).  
 FT CONFLICT 3038 3038 S -> G (IN REF. 1).  
 FT SEQUENCE 3503 AA; 379774 MW; 975B09F059F7EEF5 CRC64;  
 Query Match 55.9%; Score 38; DB 1; Length 3503;  
 Best Local Similarity 60.0%; Pred. No. 1.6e+02;  
 Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

```

QY      1 DVPYPTSY 10
SVR_BIFLO
Db      149 DLEPTQRY 158

RESULT 11
SVR_BIFLO
ID      SYR_BIFLO      STANDARD;      PRT;      620 AA.
AC      O8GV2;
DT      15-MAR-2004 (Rel. 43, Created)
DT      15-MAR-2004 (Rel. 43, Last sequence update)
DT      15-MAR-2004 (Rel. 43, Last annotation update)
DE      Arginyl-tRNA synthetase (EC 6.1.1.19) (Arginine--tRNA ligase) (ARGRS).
GN      ARGS OR B11272.
OS      Bifidobacterium longum.
OC      Bacteria; Actinobacteria; Actinobacteriales; Bifidobacteriales;
OC      Bifidobacteriaceae; Bifidobacterium.
OX      NCBI_TaxID=216816;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=NCC 2705;
RX      MEDLINE=22294977; PubMed=12381787;
RA      Schell M.A., Karimkhanlou M., Snel B., Vilanova D., Berger B.,
RA      Pessi G., Zwahlen M.-C., Desiere F., Bork P., Delley M.,
RA      Fridmore R.D., Arigoni F.;
RT      "The genome sequence of Bifidobacterium longum reflects its adaptation
RT      to the human gastrointestinal tract.";
RL      Proc. Natl. Acad. Sci. U.S.A. 99:14422-14427(2002).
CC      -!- CATALYTIC ACTIVITY: ATP + L-arginine + tRNA(Arg) = AMP +
CC      diphosphate + L-arginyl-tRNA(Arg).
CC      -!- SUBUNIT: Monomer (By similarity).
CC      -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC      -!- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
DR      EMBL: AE014754; AAN25073.1; --
DR      HAVAP; MF 00123; 1
DR      InterPro; IPR001278; Arg_tRNA-synt_1c.
DR      InterPro; IPR005148; N.
DR      InterPro; IPR008909; tRNA-synt_1d_C.
DR      InterPro; IPR001412; tRNA-synt_1.
DR      InterPro; IPR009080; tRNA-synt_1a_bind.
DR      Pfam; PF03485; N-Arg; 1.
DR      Pfam; PF00750; tRNA-synt_1d; 1.
DR      Pfam; PF05746; tRNA-synt_1c; 1.
DR      PRINTS; PR01038; TRNASYNTHARG.
DR      TIGRFAMs; TIGR00456; args; 1.
DR      PROSITE; PS00178; AA tRNA LIGASE I; 1.
KW      Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW      Complete proteome.
FT      SITE 147 "HIGH" REGION
SQ      SEQUENCE 620 AA; 67206 MW; F51C3C0C8CD9AB06 CRC64;

Query Match 54.4%; Score 37; DB 1; Length 620;
Best Local Similarity 50.0%; Pred. No. 37;
Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY      1 DVPYPTSY 12
Db      534 DRQPHKVARYLE 545

RESULT 12
DEFC_AEDAE
ID      DEFC_AEDAE      STANDARD;      PRT;      99 AA.

```

```

AC      P81603; Q9YOF0; Q9YOF1;
DT      30-MAY-2000 (Rel. 39, Created)
DT      18-OCT-2001 (Rel. 40, Last sequence update)
DT      10-OCT-2003 (Rel. 42, Last annotation update)
DE      Defensin C precursor.
OS      Aedes aegypti (Yellowfever mosquito).
OC      Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC      Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Aedes.
OX      NCBI_TaxID=7159;
RN      [1]
RP      SEQUENCE FROM N.A., DEVELOPMENTAL STAGE, INDUCTION, AND SUBCELLULAR
RP      LOCATION.
RC      STRAIN=Liverpool; TISSUE=Fat body;
RX      MEDLINE=99124369; PubMed=9927179;
RA      Lowenberger C.A., Smartt C.T., Bulet P., Ferdig M.T., Severson D.W.,
RA      Hoffmann J.A., Christensen B.M.;
RT      "Insect immunity: isolation of three novel inducible antibacterial
RT      defensins from the vector mosquito, Aedes aegypti.";
RL      Insect Biochem. Mol. Biol. 25:867-873(1995).
RN      [3]
RP      SEQUENCE OF 60-99.
RC      STRAIN=REFM;
RX      MEDLINE=96047965; PubMed=7568275;
RA      Chalk R., Albuquerque C.M., Ham P.J., Townson H.;
RT      "Full sequence and characterization of two insect defensins: immune
RT      peptides from the mosquito Aedes aegypti.";
RL      Proc. R. Soc. Lond., B, Biol. Sci. 261:217-221(1995).
CC      -!- FUNCTION: Antibacterial peptide mostly active against Gram-
CC      positive bacteria.
CC      -!- SUBCELLULAR LOCATION: Secreted.
CC      -!- TISSUE SPECIFICITY: Hemolymph.
CC      -!- DEVELOPMENTAL STAGE: Expressed 30 minutes after infection and
CC      remained present through to 21 days. Expressed in white or callow
CC      pupae during metamorphosis, but no expression was seen in larvae.
CC      -!- INDUCTION: By bacterial infection.
CC      -!- POLYMORPHISM: There are two defensin C isoforms, C1 (shown here)
CC      and C2.
CC      -!- SIMILARITY: Belongs to the arthropod defensin family. Subfamily 1.
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
DR      EMBL: AF156092; AAD40116.2; --
DR      EMBL; AF156093; AAD40117.2; --
DR      HSSP; P10891; IICA.
DR      InterPro; IPR001542; Defensin_anpod.
DR      InterPro; IPR003614; Knot1.
DR      Pfam; PF01097; Arthro defensin; 1.
DR      PRINTS; PR00271; DEFENSIN.
DR      SMART; SM00505; Knot1.
DR      PROSITE; PS00425; ARTHROPOD DEFENSINS; 1.
KW      Insect immunity; Antibiotic; POTENTIAL.
FT      SIGNAL 1 23
FT      PROPEP 24 59
FT      CHAIN 60 99
FT      BY SIMILARITY.
FT      DISULFID 62 89
FT      BY SIMILARITY.
FT      DISULFID 75 95
FT      BY SIMILARITY.
FT      DISULFID 79 97

```

```

FT METAL 53 IRON (BY SIMILARITY).
FT METAL 57 IRON (BY SIMILARITY).
FT METAL 124 IRON (BY SIMILARITY).
SQ SEQUENCE 157 AA; 18098 MW; 89C3844C37BB7504 CRC64;

Query Match 52.9%; Score 36; DB 1; Length 157;
Best Local Similarity 46.2%; Pred. No. 13;
Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 DVRPYETSRYLEI 13
   | : | : | : |
Db 133 DMAKVESSKYLQI 145

RESULT 14
DHAA_MYCBO STANDARD; PRT; 300 AA.
ID DHAA_MYCBO
AC Q9XB14;
DT 28-FEB-2003 (Rel. 41, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Haloalkane dehalogenase [EC 3.8.1.5].
GN DHAAC OR MB2610.
OS Mycobacterium bovis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
ON NCBI_TaxID=1765;
RX [1]
RN SEQUENCE FROM N.A.
RP SEQUENCE FROM N.A.
RC STRAIN=M011;
RA Jesenska A., Rychlik I., Pavlik I., Damborsky J.;
RT "Distribution of haloalkane dehalogenase genes among bovine
RT mycobacteria.";
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=Ar2122/97;
RC MEDLINE=22709107; PubMed=12788972;
RA Garnier T., Eiglmair K., Camus J.-C., Medina N., Mansoor H.,
RA Pryor M., Duthoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,
RA Harris B., Atkin R., Doggett J., Maves R., Keating L., Wheeler P.R.,
RA Packhill J., Barrall B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
RT "The complete genome sequence of Mycobacterium bovis.";
RL Proc. Natl. Acad. Sci. U.S.A. 190:7877-7882(2003).
CC -I- FUNCTION: Catalyzes hydrolytic cleavage of carbon-halogen bonds in
CC halogenated aliphatic compounds, leading to the formation of the
CC corresponding primary alcohols, halide ions and protons (By
CC similarity).
CC -I- CATALYTIC ACTIVITY: 1-haloalkane + H(2)O = a primary alcohol +
CC halide.
CC -I- SUBUNIT: Monomer (By similarity).
CC -I- SIMILARITY: Belongs to the haloalkane dehalogenase family.
CC Subfamily 2.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on use
CC by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; A243259; CAB45532.1; -
CC EMBL; BX248343; CAD94795.1; -
CC HAMAP; MF_01231; -; 1.
CC InterPro; IPR0000073; A/b hydrolase.
CC InterPro; IPR000639; Epox hydrolase.
CC InterPro; IPR000379; Ser esters.
CC Pfam; PF00561; abhydrolase_1.
CC PRINTS; PR00412; EPOXYDRLAS.
CC Hydrolase; Complete proteome.
CC -----
CC DR EMBL; A243259; CAB45532.1; -
CC DR EMBL; BX248343; CAD94795.1; -
CC DR HAMAP; MF_01231; -; 1.
CC DR InterPro; IPR0000073; A/b hydrolase.
CC DR InterPro; IPR000639; Epox hydrolase.
CC DR InterPro; IPR000379; Ser esters.
CC DR Pfam; PF00561; abhydrolase_1.
CC DR PRINTS; PR00412; EPOXYDRLAS.
CC DR Hydrolase; Complete proteome.
CC -----
CC KW Hydrolase; Complete proteome.
CC NUCLEOPHILE (BY SIMILARITY).
CC FT ACT SITE 109 109
CC FT ACT SITE 133 133
CC FT ACT SITE 133 133
CC ACID (BY SIMILARITY).

```

DR EMEL; ABO07099; AAK46969.1; -  
DR PIR; B70725; B70725.  
DR TIGR; MT2656; -  
DR TUBERCULIST; RV2579; -  
DR HAMAP; MF\_01231; -; 1.  
DR InterPro; IPR000073; A/b hydrolase.  
DR InterPro; IPR000639; Epox hydrolase.  
DR InterPro; IPR000379; Ser esters.  
DR Pfam; PF00561; abhydrolase\_1.  
DR PRINTS; PR00412; EPOXYHYDROLASE.  
DR Hydrolase; Complete proteome.  
KW Hydrolase; Complete proteome.  
FT ACT SITE 109 109 NUCLEOPHILE (BY SIMILARITY).  
FT ACT SITE 133 133 ACID (BY SIMILARITY).  
FT ACT SITE 273 273 BASE (BY SIMILARITY).  
SQ SEQUENCE 300 AA; 33728 MW; E8F4854749F22562 CRC64;  
Query Match 52.9%; Score 36; DB 1; Length 300;  
Best Local Similarity 58.3%; Pred. No. 26;  
Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;  
QY 2 VRPYRTSRYLEI 13  
DB 6 VEPYGPQKYLEI 17

Search completed: October 5, 2004, 08:01:39  
Job time: 2.16358 secs

FT ACT SITE 273 273 BASE (BY SIMILARITY).  
FT CONFLICT 96 96 T -> A (IN REF. 1).  
FT CONFLICT 120 120 N -> K (IN REF. 1).  
SQ SEQUENCE 300 AA; 33730 MW; BCF471EC4359F7B CRC64;  
Query Match 52.9%; Score 36; DB 1; Length 300;  
Best Local Similarity 58.3%; Pred. No. 26;  
Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;  
QY 2 VRPYRTSRYLEI 13  
DB 6 VEPYGPQKYLEI 17

RESULT 15

DEHA\_MYCTU STANDARD; PRT; 300 AA.  
AC Q50642;  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Haloalkane dehalogenase 3 (EC 3.8.1.5).  
GN DEHA OR RV2579 OR MT2656 OR WTCY227.22C.  
OS Mycobacterium tuberculosis.  
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
OX NCBI\_TaxID=1773;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=H37RV;  
RX MEDLINE=98295987; PubMed=98344230;  
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,  
Baddock K., Basham D., Brown D., Chillingworth T., Connor R.,  
Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,  
Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
Rutter S., Seeger K., Skelton S., Squares S., Squares R.,  
Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;  
RT "Deciphering the biology of Mycobacterium tuberculosis from the  
complete genome sequence.";  
RL Nature 393:537-544(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX STRAIN=CDC 1551 / Oshkosh;  
RX MEDLINE=22206494; PubMed=12218036;  
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,  
Kolony J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,  
Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,  
Bisnai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;  
RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and  
laboratory strains.";  
RL J. Bacteriol. 184:5479-5490(2002).  
CC -!- FUNCTION: Catalyzes hydrolytic cleavage of carbon-halogen bonds in  
halogenated aliphatic compounds, leading to the formation of the  
corresponding primary alcohols, halide ions and protons (By  
similarity).  
CC -!- CATALYTIC ACTIVITY: 1-haloalkane + H(2)O = a primary alcohol +  
halide.  
CC -!- SUBUNIT: Monomer (By similarity).  
CC -!- SIMILARITY: Belongs to the haloalkane dehalogenase family.  
CC Subfamily 2.  
CC -----  
CC This SWISS-PROT entry is copyright.. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; 277724; CAB01264.1; -.

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2004, 07:34:11 ; Search time 6.88117 Seconds  
(without alignments)  
596.081 Million cell updates/sec

Title: US-09-805-290A-10  
Perfect score: 69  
Sequence: 1 DVRPYRTSYLEI 13

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL 25:  
1: sp archaea:  
2: sp bacteria:  
3: sp fungi:  
4: sp human:  
5: sp invertebrate:  
6: sp mammal:  
7: sp mhc:  
8: sp organelle:  
9: sp phage:  
10: sp plant:  
11: sp rodent:  
12: sp virus:  
13: sp vertebrate:  
14: sp unclassified:  
15: sp\_rvirus:  
16: sp\_bacteriap:  
17: sp\_archaeap:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	42	61.8	451	11	O35390 rattus norv
2	41	60.3	272	10	O81AQ8 arabidopsis
3	41	60.3	303	10	Q9MIK7 arabidopsis
4	41	60.3	581	10	Q9LTL1 arabidopsis
5	41	60.3	775	10	Q9SIY9 arabidopsis
6	40	58.8	442	16	Q8U8Y6 arabidopsis
7	40	58.8	462	11	Q8CON9 mus musculus
8	40	58.8	654	2	Q52001 streptococc
9	40	58.8	766	5	P91590 ciona intes
10	39	57.4	196	2	Q7W284 nonmuraea
11	39	57.4	217	5	Q25368 loligo opal
12	39	57.4	254	5	Q9VJY1 drosophila
13	39	57.4	254	5	Q8MZ39 drosophila
14	39	57.4	258	10	O80652 arabidopsis
15	39	57.4	322	16	Q7WGG8 borderella
16	39	57.4	322	16	Q7W4Z0 borderella

17	39	57.4	323	17	O8TGZ6
18	39	57.4	422	17	Q974N2
19	39	57.4	509	5	O01792
20	39	57.4	558	16	Q87BJ7 xylella fas
21	39	57.4	559	16	Q8P8P5 xanthomonas
22	39	57.4	561	16	Q8PK77 xanthomonas
23	39	57.4	569	16	Q9PAQ7 xylella fas
24	39	57.4	980	10	Q84SF5 oryza sativ
25	38	55.9	51	16	Q7UW68 rhodopirell
26	38	55.9	276	16	Q884I1 pseudomonas
27	38	55.9	415	16	Q8ZB52 salmonella
28	38	55.9	415	16	Q8BX44 escherichia
29	38	55.9	415	16	Q8FX34 escherichia
30	38	55.9	415	16	Q8Z8M7 salmonella
31	38	55.9	417	16	Q92RU9 rhizobium m
32	38	55.9	441	16	Q8X372 escherichia
33	38	55.9	726	12	Q9QLJ3 influenza b
34	38	55.9	791	5	Q8IRJ0 drosophila
35	38	55.9	856	5	Q86LP2 drosophila
36	38	55.9	952	5	Q9W0R3 drosophila
37	38	55.9	973	5	Q8WSF2 drosophila
38	38	55.9	1441	4	O60310 homo sapien
39	37	54.4	84	10	Q9LN47 arabidopsis
40	37	54.4	91	13	Q9YHE1 brachydanio
41	37	54.4	136	16	Q8PF48 xanthomonas
42	37	54.4	197	2	Q84AM7 streptomyce
43	37	54.4	221	16	Q818M9 bacillus ce
44	37	54.4	238	5	Q22600 caenorhabdi
45	37	54.4	250	10	Q8LRE0 oryza sativ

## ALIGNMENTS

RESULT 1  
O35390 PRELIMINARY; PRT; 451 AA.  
AC O35390;  
DC O35390; (TREMELrel. 05, Created)  
DT 01-JAN-1998 (TREMELrel. 05, Last sequence update)  
DE 01-NOV-1998 (TREMELrel. 08, Last annotation update)  
DE Endo-alpha-D-mannosidase.  
GN ENMAN.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Sprague-Dawley;  
RX MEDLINE=98030622; PubMed=9361017;  
RA Spiro M.J., Bhoyroo V.D., Spiro R.G.;  
RT "Molecular cloning and expression of rat liver endo-alpha-mannosidase,  
an N-linked oligosaccharide processing enzyme.";  
RL J. Biol. Chem. 272:29356-29363(1997).  
DR EMBL; AF023657; AAB86925.1; -.  
SQ SEQUENCE 451 AA; 51671 MW; 5A987BFD07585066 CRC64;

Query Match 61.8%; Score 42; DB 11; Length 451;  
Best Local Similarity 61.5%; Pred. No. 22;  
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 DVRPYRTSYLEI 13  
| | | | |  
Db 415 DYRPHKPSLYLEI 427

RESULT 2  
Q81AQ8 PRELIMINARY; PRT; 272 AA.  
ID Q81AQ8;  
AC Q81AQ8;  
DT 01-OCT-2002 (TREMELrel. 22, Created)  
DT 01-OCT-2002 (TREMELrel. 22, Last sequence update)

```

DB 10 DIRPKTSWRIQV 22

RESULT 4
Q9LTL1 PRELIMINARY; PRT; 581 AA.
ID Q9LTL1;
AC Q9LTL1;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Replication protein A1-like.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Sato S., Nakamura Y., Kaneko T., Kato T., Asamizu E., Tabata S.;
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
[2]
RN SEQUENCE FROM N.A.
RP STRAIN=Columbia;
RX MEDLINE=20277480; PubMed=10819329;
RA Nakamura Y.;
RT "Structural analysis of Arabidopsis thaliana chromosome 3. I. Sequence
RT features of the regions of 4,504,864 bp covered by sixty P1 and TAC
RT clones."
RL DNA Res. 7:131-135(2000).
DR EMBL; AB024034; BAB02796.1; -.
DR InterPro; IPR003871; DUF223.
DR InterPro; IPR008994; Nucleic_acid_OB.
DR Pfam; PF02721; DUF223; 1.
SQ SEQUENCE 581 AA; 64867 MW; 5BCB4DB5B0E006DC CRC64;

Query Match 60.3%; Score 41; DB 10; Length 581;
Best Local Similarity 53.8%; Pred. No. 44;
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 DVPRTSRYLEI 13
DB 10 DVPRTSRVQV 22

RESULT 5
Q9SIY9 PRELIMINARY; PRT; 775 AA.
ID Q9SIY9;
AC Q9SIY9;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Putative W4-40 repeat protein.
GN AT2G40360.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RX MEDLINE=20083487; PubMed=10617197;
RA Fujii C.Y., Mason I.M., Bowman C.L., Barnstead M.E., Feldbrum T.V.,
RA Buell C.R., Ketchum K.A., Lee J.J., Renning C.M., Koo H., Moffat K.S.,
RA Cronin L.A., Shen M., VanAken S.E., Umayam L., Tallon L.J., Gill J.E.,
RA Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
RA Copenhaver G.P., Preuss D., Nierman W.C., White O., Eissen J.A.,
RA Salzberg S.L., Fraser C.M., Venter J.C.;
RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis
RT thaliana."
RL Nature 402:761-768(1999).

```

```

DB 10 DIRPKTSWRIQV 22

RESULT 4
Q9LTL1 PRELIMINARY; PRT; 581 AA.
ID Q9LTL1;
AC Q9LTL1;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Replication protein A1-like.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Sato S., Nakamura Y., Kaneko T., Kato T., Asamizu E., Tabata S.;
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
[2]
RN SEQUENCE FROM N.A.
RP STRAIN=Columbia;
RX MEDLINE=20277480; PubMed=10819329;
RA Nakamura Y.;
RT "Structural analysis of Arabidopsis thaliana chromosome 3. I. Sequence
RT features of the regions of 4,504,864 bp covered by sixty P1 and TAC
RT clones."
RL DNA Res. 7:131-135(2000).
DR EMBL; AB024034; BAB02796.1; -.
DR InterPro; IPR003871; DUF223.
DR InterPro; IPR008994; Nucleic_acid_OB.
DR Pfam; PF02721; DUF223; 1.
SQ SEQUENCE 581 AA; 64867 MW; 5BCB4DB5B0E006DC CRC64;

Query Match 60.3%; Score 41; DB 10; Length 581;
Best Local Similarity 53.8%; Pred. No. 44;
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 DVPRTSRYLEI 13
DB 10 DVPRTSRVQV 22

RESULT 5
Q9SIY9 PRELIMINARY; PRT; 775 AA.
ID Q9SIY9;
AC Q9SIY9;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Putative W4-40 repeat protein.
GN AT2G40360.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RX MEDLINE=20083487; PubMed=10617197;
RA Fujii C.Y., Mason I.M., Bowman C.L., Barnstead M.E., Feldbrum T.V.,
RA Buell C.R., Ketchum K.A., Lee J.J., Renning C.M., Koo H., Moffat K.S.,
RA Cronin L.A., Shen M., VanAken S.E., Umayam L., Tallon L.J., Gill J.E.,
RA Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
RA Copenhaver G.P., Preuss D., Nierman W.C., White O., Eissen J.A.,
RA Salzberg S.L., Fraser C.M., Venter J.C.;
RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis
RT thaliana."
RL Nature 402:761-768(1999).

```



```

DR InterPro; IPR000205; NAD_BS.
DR Pfam; PF01266; DAO; 1.
KW Complete proteome.
SQ SEQUENCE 442 AA; 48309 MW; E074DB4BCFFBB62D CRC64;
    Query Match          58.8%; Score 40; DB 16; Length 442;
    Best Local Similarity 58.3%; Pred. No. 51;
    Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 DVRPYRTSYLEI 12
   | | | | | | | |
Db 421 DPRPYRSRFFD 432

RESULT 7
ID Q8CON9 PRELIMINARY; PRT; 462 AA.
AC Q8CON9;
DT 01-NAR-2003 (TrEMBLrel. 23, Created)
DT 01-NAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Endo-alpha-D-mannosidase homolog.
GN 4932703L02Rik.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]_RN
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/60; TISSUE=Testis;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium.
RA The RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573 (2002).
DR EMBL; AK030141; BAC28605.1; -
DR MGI; MGI:2444484; 4932703L02Rik.
SQ SEQUENCE 462 AA; 53199 MW; 5EA81B9C5075E684 CRC64;
    Query Match          58.8%; Score 40; DB 11; Length 462;
    Best Local Similarity 53.8%; Pred. No. 54;
    Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 DVRPYRTSYLEI 13
   | | | | | | |
Db 426 DYRPHKPSLYLEL 438

RESULT 8
ID Q52001 PRELIMINARY; PRT; 654 AA.
AC Q52001;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE NICKASE.
OS Streptococcus sp., and
OS Enterococcus faecalis (Streptococcus faecalis).
CG Plasmid pIP501, and plasmid pRE25.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1306, 1351;
RN [1]_RN
RP SEQUENCE FROM N.A.
RC SPECIES=Streptococcus sp.; STRAIN=EVA1702; PLASMID=pIP501;
RX MEDLINE=95362640; PubMed=7635806;
RA Wang A., Macrina F.L.;
RT "Streptococcal plasmid pIP501 has a functional orit site.";
RL J. Bacteriol. 177:4199-4206 (1995).
RN [2]_RN
RP SEQUENCE FROM N.A.
RC SPECIES=E.faecalis; STRAIN=RE25; PLASMID=pRE25;

```

RX PubMed=11735367;  
 RA Schwarz F.V., Perreten V., Teuber M.;  
 RT "Sequence of the 50-kb conjugative multiresistance plasmid PRE25 from  
 RL Enterococcus faecalis R825.";  
 DR EMBL; L39769; AAA99466.1; -;  
 DR EMBL; X92945; CAC29179.1; -;  
 DR GO; GO:0046821; C:extrachromosomal DNA; IEA.  
 DR InterPro; IPR005053; MobA\_MobL.  
 DR Pfam; PF03389; MobA\_MobL\_1.  
 KW PfamID.  
 SQ SEQUENCE 654 AA; 76493 MW; C3EE58B8B4468600 CRC64;

Query Match 58.8%; Score 40; DB 2; Length 654;  
 Best Local Similarity 60.0%; Pred. No. 77;  
 Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 3 RPYRTSRYLE 12  
 :||:|||||  
 Db 418 KPFQTSRYLD 427

## RESULT 9

P91590  
 ID P91590 PRELIMINARY; PRT; 766 AA.

AC P91590;  
 DT 01-MAY-1997 (TrEMBLrel. 03, Created)  
 DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Transglutaminase.  
 OS Ciona intestinalis.  
 OC Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;  
 OC Phlebobranchia; Clonidae; Ciona.  
 OX NCBI\_TaxID=7719;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=9733026; PubMed=9187361;  
 RA Carliello L., Ristoreto F., Zanetti L.;  
 RT "A new transglutaminase-like from the ascidian Ciona intestinalis.";  
 RL FEBS Lett. 408:171-176(1997).  
 RN [2]

RP SEQUENCE FROM N.A.  
 RA Carliello L.;  
 RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.

DR EMBL; Y10212; CAAV1283.1; -;  
 DR HSSP; P00488; 1GGU.  
 DR InterPro; IPR001102; GluttransfG.  
 DR InterPro; IPR007110; IG-like.  
 DR InterPro; IPR008958; Transglut C.  
 DR InterPro; IPR002931; Transglutase-like.  
 DR Pfam; PF00927; Transglutamin C/I.  
 DR Pfam; PF00868; Transglutamin N/I.  
 DR Pfam; PF01841; Transglut\_core; 1.  
 DR SMART; SM00460; TGC; 1.  
 SQ SEQUENCE 766 AA; 87194 MW; 7F90D921D87FDD1D CRC64;

Query Match 58.8%; Score 40; DB 5; Length 766;  
 Best Local Similarity 77.8%; Pred. No. 90;  
 Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 DVRPYRTSR 9  
 :|||||  
 Db 732 EVRPYRSSR 740

## RESULT 10

Q7WZ84  
 ID Q7WZ84 PRELIMINARY; PRT; 196 AA.

AC Q7WZ84;  
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Putative Vary-type carboxypeptidase.  
 GN DBV.  
 OS Nonomurea sp. ATCC 39727.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Streptosporangineae; Streptosporangiaceae; Nonomurea.  
 OX NCBI\_TaxID=93944;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 39727;  
 RA Sasio M., Stinchi S., Beltrametti P., Lazzarini A., Donadio S.;  
 RT "The gene cluster for the biosynthesis of the glycopeptide antibiotic  
 RT A40926 by Nonomurea sp.";  
 RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AJ561198; CAD91202.1;  
 KW Acyltransferase; Carboxypeptidase; Glycosyltransferase; Monooxygenase.  
 SQ SEQUENCE 196 AA; 22157 MW; 003AA44D15127982 CRC64;

Query Match 57.4%; Score 39; DB 2; Length 196;  
 Best Local Similarity 58.3%; Pred. No. 34;  
 Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 DVRPYRTSRYLE 12  
 ||||:|:  
 Db 137 DVPHEGAWLE 148

## RESULT 11

Q25368  
 ID Q25368 PRELIMINARY; PRT; 217 AA.

AC Q25368;  
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE S-crystallin.  
 OS Loligo opalescens (California market squid).  
 OC Eukaryota; Metazoa; Mollusca; Cephalopoda; Colecoidea; Neocoleoidea;  
 OC Decapodiformes; Loliginidae; Loligo.  
 OX NCBI\_TaxID=31211;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX TISSUE=Lens;  
 RC MEDLINE=96139039; PubMed=8587103;  
 RA Tomarev S.I., Chung S., Piatigorsky J.;  
 RT "Glutathione S-transferase and S-crystallins of cephalopods: evolution  
 RT from active enzyme to lens-refractive proteins.";  
 RL J. Mol. Evol. 41:1048-1056(1995).

DR EMBL; U19300; AAA97551.1; -;  
 DR HSSP; P46088; 2GSQ.  
 DR InterPro; IPR004046; GST\_Cterm.  
 DR InterPro; IPR004045; GST\_Nterm.  
 DR InterPro; IPR003083; S-crystallin.  
 DR Pfam; PF00043; GST\_C; 1.  
 DR Pfam; PF02938; GST\_N; 1.  
 DR PRINTS; PRO1269; SCRYSTALLIN.

SQ SEQUENCE 217 AA; 26113 MW; 66D4D97D9B3B6F8A CRC64;

Query Match 57.4%; Score 39; DB 5; Length 217;  
 Best Local Similarity 50.0%; Pred. No. 38;  
 Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 DVRPYRTSRYLE 12  
 |:|:|:  
 Db 125 DLNPFYMQTRYME 136

## RESULT 12

Q9VUY1  
 ID Q9VUY1 PRELIMINARY; PRT; 254 AA.

AC Q9VUY1;  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE CG16848 protein.

CG16848.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Berkeley;  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Anantides P.G., Scher S.E., Li P.W., Hoskins R.A., Galie R.P.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.C., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotier P.,  
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Foshier C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wassarman D.A., Weissstock G.M., Weissbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zavari J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RL "the genome sequence of Drosophila melanogaster";  
RL Science 287:2185-2195(2000).  
DR EMBL; AB003640; AAF53304.1; -.  
DR FlyBase; FBgn0032522; CG16848.  
DR GO; GO:0016740; P:transferase activity; IEA.  
DR GO; GO:0008152; P:metabolism; IEA.  
DR InterPro; IPR002500; PAPS\_reduct.  
DR Pfam; PF01507; PAPS\_reduct; 1.  
SQ SEQUENCE 254 AA; 29723 MW; 3D5164772719F06F CRC64;  
Query Match 57.4%; Score 39; DB 5; Length 254;  
Best Local Similarity 72.7%; Pred. No. 44;  
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
Qy 3 RPYRTSRYLEI 13  
Db 13 RHYRTSRWLHI 23  
RESULT 13  
ID Q8MZ99 PRELIMINARY; PRT; 254 AA.  
AC Q8MZ99;  
DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE AT21573p.

CG16848.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,  
RA Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,  
RA George R., Gonzalez M., Guarin H., Kronmiller B., Li P., Liao G.,  
RA Miranda A., Mungall C.J., Nunco J., Pacleb J., Paragas V., Park S.,  
RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,  
RA Celniker S.;  
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AY113289; AAM29294.1; -.  
DR FlyBase; FBgn0032522; CG16848.  
DR GO; GO:0016740; P:transferase activity; IEA.  
DR GO; GO:0008152; P:metabolism; IEA.  
DR InterPro; IPR002500; PAPS\_reduct.  
DR Pfam; PF01507; PAPS\_reduct; 1.  
SQ SEQUENCE 254 AA; 37F06F7895A282C6 CRC64;  
Query Match 57.4%; Score 39; DB 5; Length 254;  
Best Local Similarity 72.7%; Pred. No. 44;  
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
Qy 3 RPYRTSRYLEI 13  
Db 13 RHYRTSRWLHI 23  
RESULT 14  
ID O80652 PRELIMINARY; PRT; 258 AA.  
AC O80652;  
DT 01-NOV-1998 (TrEMBLrel. 08, Created)  
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE T14N5.4 protein.  
GN T14N5.4.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Columbia;  
RA Federspiel N.A., Palm C.J., Conway A.B., Kurtz D.B., Conway A.R.,  
RA Au M., Araujo R., Buehler E., Dewar K., Feng J., Kim C., Li Y.,  
RA Oji O., Osborne B.I., Shinn P., Sun H., Toriumi M., Vysotskaia V.S.,  
RA Yu G., Ecker J., Theologis A., Davis R.W.;  
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AC004260; AAC34353.1; -.  
DR PIR; T00447; T00447.  
DR InterPro; IPR006502; DUF\_plant\_1615.  
DR Pfam; PF04720; DUF506; 1.  
DR TIGRFAMs; TIGR01615; A.thal.3542; 1.  
SQ SEQUENCE 258 AA; 29578 MW; FBC341434DFA6380 CRC64;  
Query Match 57.4%; Score 39; DB 10; Length 258;  
Best Local Similarity 66.7%; Pred. No. 45;  
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;  
Qy 4 PYRTSRYLEI 12  
Db 199 PYRTSRVWQ 207  
RESULT 15  
ID Q7WGG8 PRELIMINARY; PRT; 322 AA.  
Q7WGG8

AC Q7WCG8;  
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Putative transcriptional regulator.  
 GN BB3951.  
 OS Bordetella bronchiseptica (Alcaligenes bronchisepticus).  
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;  
 OC Alcaligenaceae; Bordetella.  
 OX NCBI\_TaxID=518;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=RB50 / ATCC BAA-588;  
 RX MEDLINE=2827954; PubMed=12910271;  
 RA Parkhill J., Sebahia M., Preston A., Murphy L.D., Thomson N.,  
 RA Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,  
 RA Cerdeno-Tarraga A.M., Temple L., James K., Harris B., Quail M.A.,  
 RA Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,  
 RA Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,  
 RA Feltwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jagels K.,  
 RA Leather S., Meule S., Norberczak H., O'Neil S., Ormond D., Price C.,  
 RA Rabinowitch E., Rutter S., Sanders M., Saunders D., Seeger K., Stevens K.,  
 RA Sharp S., Simmonds M., Skelton J., Squares R., Squares S.,  
 RA Unwin L., Whitehead S., Barrell B.G., Maskell D.J.;  
 RT "Comparative analysis of the genome sequences of Bordetella pertussis,  
 RT Bordetella parapertussis and Bordetella bronchiseptica.";  
 RL Nat. Genet. 35:32-40(2003).  
 DR EMBL, BX640449; CAE34314.1; --  
 KW Complete proteome.  
 SQ SEQUENCE 322 AA; 34669 MW; F6072C095666D3E6 CRC64;  
 Query Match 57.4%; Score 39; DB 16; Length 322;  
 Best Local Similarity 53.8%; Pred. No. 57;  
 Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;  
 QY 1 DVPYRTSHYLEI 13  
 DB 170 DVTYRTDLVAL 182

Search completed: October 5, 2004, 08:13:14  
 Job time : 7.88117 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2004, 07:04:10 ; Search time 9.89043 Seconds  
(without alignments)  
371.381 Million cell updates/

Title: US-09-805-290A-10

Perfect score:

Sequence: 1·DVRPYRTSRYLEI 1:

Scoring table: BLOSUM62

Gapo 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length:	2000000000
Maximum DB seq length:	2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A Geneseq 29Jan04:\*

- 1: geneseqp1980s: \*  
2: geneseqp1990s: \*  
3: geneseqp2000s: \*  
4: geneseqp2001s: \*  
5: geneseqp2002s: \*  
6: geneseqp2003as: \*  
7: geneseqp2003bs: \*  
8: geneseqp2004s: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	68	100.0	13	4	AAE10542	Llama spe
2	68	100.0	130	4	AAE10559	HPL inhib
3	67	98.5	13	4	AAE10540	Llama spe
4	67	98.5	130	4	AAE10552	HPL inhib
5	66	97.1	13	4	AAE10541	Llama spe
6	66	97.1	130	4	AAE10554	HPL inhib
7	64	94.1	13	4	AAE10543	Llama spe
8	42	61.8	142	3	AAE45037	Human sec
9	42	61.8	451	7	ADP54766	Rat Prote
10	41	60.3	252	3	AAQ06656	Arabidops
11	41	60.3	272	3	AAQ06655	Arabidops
12	41	60.3	281	3	AAQ06654	Arabidops
13	41	60.3	727	6	ABU27962	Protein e
14	41	60.3	944	6	ABU45234	Protein e
15	41	60.3	961	6	ABU32202	Protein e
16	39	57.4	90	6	ABP75655	Human sec
17	39	57.4	208	3	AAQ31539	Arabidops
18	39	57.4	238	3	AAQ31528	Arabidops
19	39	57.4	250	3	AAV73330	HTM clon
20	39	57.4	254	4	ABY70887	Drosophi
21	39	57.4	268	3	AAQ31527	Arabidops
22	39	57.4	509	5	ABQ30966	Herbicida
23	39	57.4	558	6	ABU33563	Protein e
24	38	55.9	52	4	AAU57763	Propioni
25	38	55.9	52	6	ABM54282	Propioni

## ALIGNMENTS

RESULT 1  
AAE10542  
ID AAE10542 standard; peptide; 13 AA.

AC AAE10542;

DT 10-DEC-2001 (first entry)

DE Llama species antibody VHH CDR3 #11.

llama antibody; camelid; anorectic; heavy chain variable domain; VHH; human dietary enzyme inhibitor; medication; human pancreatic lipase; HPU; food; human gastric lipase; HGL; cosmetic control; body weight; complementarity determining region; CDR3.

OS Lama sp.

PN EP1134231-A1.

XX  
PD  
19-SEP-2001.

20-FEB-2001: 2001EP-00200703.

14-MAR-2000: 2000EP-00200930.

PA (UNIL ) UNILEVER NV.

PA (UNIL ) UNILEVER PLC.

PI Bezemer S, Van De Burg M, De Haard JJW, Tareilus E;

WPI: 2001-572718/65.

24 New antibody or its fragments for inhibiting human dietary enzymes,  
PT  
PT useful for cosmetic control of body weight of human beings, comprises  
PT heavy chain variable domain derived from immunoglobulin naturally devoid  
PT of light chains.

PS Disclosure: Page 17: 37pp: English.

The patent discloses antibodies or their fragments comprising a heavy chain variable domain (VHH) derived from an immunoglobulin naturally devoid of light chains specific for inhibiting human dietary enzymes. The antibodies of the invention are useful for the preparation of medicaments or food for inhibiting the activity of one or more human dietary enzymes especially human pancreatic lipase (HPL) or human gastric lipase (HGL) which are useful for the cosmetic control of body weight of human beings. The present peptide sequence is a complementary determining region 3 (CDR3) of llama species (camelid) antibody VHH region

```

SQ      Sequence 13 AA;
Query Match      100.0%; Score 68; DB 4; Length 13;
Best Local Similarity 100.0%; Pred. No. 5.8e-06;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 DVRPYRTSRYLEI 13
Db      1 DVRPYRTSRYLEI 13

RESULT 2
AAE10559
ID AAE10559 standard; peptide; 130 AA.
AC AAE10559;
XX
XX
DT 10-DEC-2001 (first entry)
DE HPL inhibiting VHH fragment, HPL #30 from llama species.
XX
XX Llama antibody; camelid; anorectic; heavy chain variable domain; VHH;
XX human dietary enzyme inhibitor; medicament; human pancreatic lipase; HPL;
XX food; human gastric lipase; HGL; cosmetic control; body weight.
XX
XX Lama sp.
XX
XX Key Location/Qualifiers
FH Region 31..35
FT /label= CDR1
FT /note= "Complementarity determining region 1"
FT Region 49..64
FT /label= CDR2
FT /note= "Complementarity determining region 2"
FT Region 98..110
FT /label= CDR3
FT /note= "Complementarity determining region 3"
XX
XX EPI134231-A1.
XX
XX 19-SEP-2001.
XX
XX 20-FEB-2001; 2001EP-00200703.
XX
XX 14-MAR-2000; 2000EP-00200930.
XX
XX (UNIL ) UNILEVER NV.
XX (UNIL ) UNILEVER PLC.
XX
XX Bezemer S, Van De Burg M, De Haard JJW, Tareilus E;
XX WPI; 2001-572718/65.
XX
XX New antibody or its fragments for inhibiting human dietary enzymes,
XX useful for cosmetic control of body weight of human beings, comprises
XX heavy chain variable domain derived from immunoglobulin naturally devoid
XX of light chains.
XX
XX Disclosure; Page 17; 37pp; English.
XX
XX The patent discloses antibodies or their fragments comprising a heavy
XX chain variable domain (VHH) derived from an immunoglobulin naturally
XX devoid of light chains specific for inhibiting human dietary enzymes. The
XX antibodies of the invention are useful for the preparation of medicaments
XX or food for inhibiting the activity of one or more human dietary enzymes
XX especially human pancreatic lipase (HPL) or human gastric lipase (HGL)
XX which are useful for the cosmetic control of body weight of human beings.
XX The present peptide sequence is HPL inhibiting VHH fragment, HPL #30 from
XX llama (camelid) species
XX
XX Sequence 130 AA;

Query Match      98.5%; Score 67; DB 4; Length 13;
Best Local Similarity 92.3%; Pred. No. 8.9e-06;
Matches 12; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      1 DVRPYRTSRYLEI 13
Db      1 DVRPYRTSRYLEV 13

RESULT 4
AAE10552
ID AAE10552 standard; peptide; 130 AA.
XX

```







XX WPI; 2000-611704/58.  
XX Nucleic acid molecules encoding human secreted proteins, used in  
PT preventing, treating or ameliorating a disorder, e.g. Alzheimer's and  
PT Parkinson's diseases and cancers.  
XX Disclosure; Page 404-405; 418pp; English.  
XX This invention describes novel isolated nucleic acid molecules (I)  
CC encoding a human secreted proteins (II) which have immunosuppressive,  
CC antiarthritic, antirheumatic, antiproliferative, cytostatic, cardiant,  
CC vasotropic, cerebroprotective, neurotropic, neuroprotective, antibacterial,  
CC virucide, fungicide, ophthalmological and vulvovaginal activity and can be  
CC used for gene therapy. (I) and (II) are used to prevent, treat or  
CC ameliorate a medical condition in e.g. humans, mice, rabbits, goats,  
CC horses, cats, dogs, chickens or sheep. (I) and (II) are also used in  
CC diagnosing a pathological condition or susceptibility to a pathological  
CC condition. The antibodies to (II) can also be used in alleviating  
CC symptoms associated with the disorders and in diagnostic immunoassays  
CC e.g. radioimmunoassays or enzyme linked immunosorbent assays (ELISA).  
CC Disorders which are diagnosed or treated include autoimmune diseases e.g.  
CC rheumatoid arthritis, hyperproliferative disorders e.g. neoplasms of the  
CC breast or liver, cardiovascular disorders e.g. cardiac arrest,  
CC cerebrovascular disorders e.g. cerebral ischemia, angiogenesis, nervous  
CC system disorders e.g. Alzheimer's disease, infections caused by bacteria,  
CC viruses and fungi and ocular disorders e.g. corneal infection. The  
CC polypeptides can also be used to aid wound healing and epithelial cell  
CC proliferation, to prevent skin aging due to sunburn, to maintain organs  
CC before transplantation, for supporting cell culture of primary tissues,  
CC to regenerate tissues and in chemotaxis. The polypeptides can also be  
CC used as a food additive or preservative to increase or decrease storage  
CC capabilities  
XX Sequence 142 AA;  
SQ Query Match 61.8%; Score 42; DB 3; Length 142;  
Best Local Similarity 61.5%; Pred. No. 7.2;  
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;  
QY 1 DVPYRTSRYLEI 13  
Db 106 DYRPHKPSLYLEI 118  
RESULT 9  
ADE54766  
ID ADE54766 standard; protein; 451 AA.  
XX ADE54766;  
XX 29-JAN-2004 (first entry)  
XX Rat Protein AAB86925, SEQ ID NC 571.  
DE DE  
XX Rat; pain; neuronal tissue; gene therapy; spinal segmental nerve injury;  
XX chronic constriction injury; CCI; spared nerve injury; SNI; Chung.  
OS Rattus norvegicus.  
XX WO2003016475-A2.  
PN 27-FEB-2003.  
PD 14-AUG-2002; 2002MO-US025765.  
PF 14-AUG-2001; 2001US-0312147P.  
PR 01-NOV-2001; 2001US-0346382P.  
PR 26-NOV-2001; 2001US-0333347P.  
XX (GEMO) GEN HOSPITAL CORP.  
PA (FARB) BAYER AG.  
XX

PI Woolf C, D'urso D, Befort K, Costigan M;  
XX WPI; 2003-268312/26.  
DR GENBANK; AAB86925.  
XX New composition comprising two or more isolated polypeptides, useful for  
PT preparing a medicament for treating pain in an animal.  
XX Claim 1; Page; 1017pp; English.  
PS The invention discloses a composition comprising two or more isolated rat  
XX or human polynucleotides or a polynucleotide which represents a fragment,  
CC derivative or allelic variation of the nucleic acid sequence. Also  
CC claimed are a vector comprising the novel polynucleotide, a host cell  
CC comprising the vector, a method for identifying a nucleotide sequence  
CC which is differentially regulated in an animal subjected to pain and a  
CC kit to perform the method; an array, a method for identifying an agent  
CC that increases or decreases the expression of the polynucleotide sequence  
CC that is differentially expressed in neuronal tissue of a first animal  
CC subjected to pain, a method for identifying a compound which regulates  
CC the expression of a polynucleotide sequence which is differentially  
CC expressed in an animal subjected to pain, a method for identifying a  
CC compound that regulates the activity of one or more of the  
CC polynucleotides, a method for producing a pharmaceutical composition, a  
CC method for identifying a compound or small molecule that regulates the  
CC activity in an animal of one or more of the polypeptides given in the  
CC specification, a method for identifying a compound useful in treating  
CC pain and a pharmaceutical composition comprising the one or more  
CC polypeptides or their antibodies. The polynucleotide or the compound that  
CC modulates its activity is useful for preparing a medicament for treating  
CC pain (e.g. spinal segmental nerve injury (SNI)) in an animal (e.g. gene  
CC injury (CCI) and spared nerve injury (SNI)) in an animal (shown in Table 2 of  
CC therapy). The sequence presented is a rat protein (shown in Table 2 of  
CC the specification) which is differentially expressed during pain. Note:  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic form directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.  
XX Sequence 451 AA;  
SQ Query Match 61.8%; Score 42; DB 7; Length 451;  
Best Local Similarity 61.5%; Pred. No. 26;  
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;  
QY 1 DVPYRTSRYLEI 13  
Db 415 DYRPHKPSLYLEI 427  
RESULT 10  
AAG06656  
ID AAG06656 standard; protein; 252 AA.  
XX AAG06656;  
XX 17-OCT-2000 (first entry)  
DT Arabidopsis thaliana protein fragment SEQ ID NO: 3507.  
DE Arabidopsis thaliana  
XX Protein identification; signal transduction pathway; metabolic pathway;  
XX hybridisation assay; genetic mapping; gene expression control; promoter;  
XX termination sequence.  
OS Arabidopsis thaliana.  
XX EP1033405-A2.  
PN 06-SEP-2000.  
PD 25-FEB-2000; 2000EP-00301439.  
PF 25-FEB-1999; 99US-0121825P.  
PR 05-MAR-1999; 99US-0123180P.  
XX

PR 09-MAR-1999; 99US-0123548P.  
PR 23-MAR-1999; 99US-0125788P.  
PR 25-MAR-1999; 99US-0126264P.  
PR 29-MAR-1999; 99US-0126785P.  
PR 01-APR-1999; 99US-0127462P.  
PR 06-APR-1999; 99US-0128234P.  
PR 08-APR-1999; 99US-0128714P.  
PR 16-APR-1999; 99US-0129845P.  
PR 19-APR-1999; 99US-0130077P.  
PR 21-APR-1999; 99US-0130449P.  
PR 23-APR-1999; 99US-0130510P.  
PR 23-APR-1999; 99US-0130891P.  
PR 28-APR-1999; 99US-0131449P.  
PR 30-APR-1999; 99US-0132048P.  
PR 30-APR-1999; 99US-0132407P.  
PR 04-MAY-1999; 99US-0132484P.  
PR 05-MAY-1999; 99US-0132485P.  
PR 06-MAY-1999; 99US-0132486P.  
PR 06-MAY-1999; 99US-0132487P.  
PR 07-MAY-1999; 99US-0132863P.  
PR 11-MAY-1999; 99US-0134256P.  
PR 14-MAY-1999; 99US-0134218P.  
PR 14-MAY-1999; 99US-0134219P.  
PR 14-MAY-1999; 99US-0134221P.  
PR 14-MAY-1999; 99US-0134370P.  
PR 18-MAY-1999; 99US-0134370P.  
PR 18-MAY-1999; 99US-0134768P.  
PR 19-MAY-1999; 99US-0134941P.  
PR 20-MAY-1999; 99US-0135124P.  
PR 21-MAY-1999; 99US-0135353P.  
PR 24-MAY-1999; 99US-0135629P.  
PR 25-MAY-1999; 99US-0136021P.  
PR 27-MAY-1999; 99US-0136392P.  
PR 28-MAY-1999; 99US-0136782P.  
PR 01-JUN-1999; 99US-0137222P.  
PR 03-JUN-1999; 99US-0137528P.  
PR 04-JUN-1999; 99US-0137503P.  
PR 07-JUN-1999; 99US-0137724P.  
PR 08-JUN-1999; 99US-0138094P.  
PR 10-JUN-1999; 99US-0138540P.  
PR 10-JUN-1999; 99US-0138847P.  
PR 14-JUN-1999; 99US-0138113P.  
PR 16-JUN-1999; 99US-0139452P.  
PR 16-JUN-1999; 99US-0139453P.  
PR 17-JUN-1999; 99US-0139492P.  
PR 18-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139460P.  
PR 18-JUN-1999; 99US-0139461P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 18-JUN-1999; 99US-0139763P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139899P.  
PR 23-JUN-1999; 99US-0140353P.  
PR 23-JUN-1999; 99US-0140354P.  
PR 24-JUN-1999; 99US-0140695P.  
PR 28-JUN-1999; 99US-0140823P.  
PR 29-JUN-1999; 99US-0140991P.  
PR 30-JUN-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 01-JUL-1999; 99US-0142154P.  
PR 02-JUL-1999; 99US-0142055P.  
PR 06-JUL-1999; 99US-0142390P.  
PR 08-JUL-1999; 99US-0142803P.  
PR 09-JUL-1999; 99US-0142920P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143542P.  
PR 14-JUL-1999; 99US-0143624P.

PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144085P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 19-JUL-1999; 99US-0144325P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 19-JUL-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144352P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 20-JUL-1999; 99US-0144884P.  
PR 21-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145085P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145192P.  
PR 23-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 23-JUL-1999; 99US-0145224P.  
PR 26-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145313P.  
PR 27-JUL-1999; 99US-0145318P.  
PR 27-JUL-1999; 99US-0145919P.  
PR 28-JUL-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 02-AUG-1999; 99US-0146389P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 04-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147192P.  
PR 05-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 06-AUG-1999; 99US-0147416P.  
PR 09-AUG-1999; 99US-0147493P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 10-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148319P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148565P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 16-AUG-1999; 99US-0148684P.  
PR 17-AUG-1999; 99US-0149368P.  
PR 17-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149426P.  
PR 20-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149929P.  
PR 23-AUG-1999; 99US-0149902P.  
PR 23-AUG-1999; 99US-0149930P.  
PR 25-AUG-1999; 99US-0150566P.  
PR 26-AUG-1999; 99US-0150884P.  
PR 27-AUG-1999; 99US-0151065P.  
PR 27-AUG-1999; 99US-0151066P.  
PR 27-AUG-1999; 99US-0151080P.  
PR 30-AUG-1999; 99US-0151303P.  
PR 31-AUG-1999; 99US-0151438P.  
PR 01-SEP-1999; 99US-0151930P.  
PR 07-SEP-1999; 99US-0152363P.  
PR 10-SEP-1999; 99US-0153070P.  
PR 13-SEP-1999; 99US-0153758P.  
PR 15-SEP-1999; 99US-0154018P.  
PR 16-SEP-1999; 99US-0154039P.  
PR 20-SEP-1999; 99US-0154779P.  
PR 22-SEP-1999; 99US-0155133P.  
PR 23-SEP-1999; 99US-0155486P.  
PR 24-SEP-1999; 99US-0155659P.  
PR 28-SEP-1999; 99US-0156458P.  
PR 29-SEP-1999; 99US-0156596P.  
PR 04-OCT-1999; 99US-0157117P.  
PR 05-OCT-1999; 99US-0157753P.

PR	06-OCT-1999;	99US-0157865P.	PR	16-APR-1999;	99US-0129845P.
PR	07-OCT-1999;	99US-0158023P.	PR	19-APR-1999;	99US-0130077P.
PR	08-OCT-1999;	99US-0158232P.	PR	21-APR-1999;	99US-0130449P.
PR	12-OCT-1999;	99US-0158363P.	PR	23-APR-1999;	99US-0130510P.
PR	13-OCT-1999;	99US-0159293P.	PR	23-APR-1999;	99US-0130991P.
PR	13-OCT-1999;	99US-0159294P.	PR	28-APR-1999;	99US-0131443P.
PR	13-OCT-1999;	99US-0159295P.	PR	30-APR-1999;	99US-0132048P.
PR	14-OCT-1999;	99US-0159329P.	PR	04-MAY-1999;	99US-0132407P.
PR	14-OCT-1999;	99US-0159330P.	PR	05-MAY-1999;	99US-0132484P.
PR	14-OCT-1999;	99US-0159331P.	PR	06-MAY-1999;	99US-0132485P.
PR	14-OCT-1999;	99US-0159637P.	PR	06-MAY-1999;	99US-0132487P.
PR	14-OCT-1999;	99US-0159638P.	PR	07-MAY-1999;	99US-0132863P.
PR	18-OCT-1999;	99US-0159584P.	PR	11-MAY-1999;	99US-0134256P.
PR	21-OCT-1999;	99US-0160741P.	PR	14-MAY-1999;	99US-0134218P.
PR	21-OCT-1999;	99US-0160767P.	PR	14-MAY-1999;	99US-0134219P.
PR	21-OCT-1999;	99US-0160768P.	PR	14-MAY-1999;	99US-0134221P.
PR	21-OCT-1999;	99US-0160770P.	PR	14-MAY-1999;	99US-0134370P.
PR	21-OCT-1999;	99US-0160814P.	PR	18-MAY-1999;	99US-0134768P.
PR	21-OCT-1999;	99US-0160815P.	PR	19-MAY-1999;	99US-0134941P.
PR	22-OCT-1999;	99US-0160980P.	PR	20-MAY-1999;	99US-0135124P.
PR	22-OCT-1999;	99US-0160981P.	PR	21-MAY-1999;	99US-0135353P.
PR	22-OCT-1999;	99US-0160989P.	PR	24-MAY-1999;	99US-0135629P.
PR	25-OCT-1999;	99US-0161404P.	PR	25-MAY-1999;	99US-0136021P.
PR	25-OCT-1999;	99US-0161405P.	PR	27-MAY-1999;	99US-0136392P.
PR	25-OCT-1999;	99US-0161406P.	PR	28-MAY-1999;	99US-0136782P.
PR	26-OCT-1999;	99US-0161359P.	PR	01-JUN-1999;	99US-0137222P.
PR	26-OCT-1999;	99US-0161360P.	PR	03-JUN-1999;	99US-0137528P.
PR	26-OCT-1999;	99US-0161361P.	PR	04-JUN-1999;	99US-0137502P.
PR	28-OCT-1999;	99US-0161920P.	PR	07-JUN-1999;	99US-0137724P.
PR	28-OCT-1999;	99US-0161920P.	PR	08-JUN-1999;	99US-0138094P.
PR	28-OCT-1999;	99US-0161933P.	PR	10-JUN-1999;	99US-0138540P.
PR	29-OCT-1999;	99US-0162142P.	PR	10-JUN-1999;	99US-0138847P.
PR	29-OCT-1999;	99US-0162142P.	PR	14-JUN-1999;	99US-0139119P.
PR	29-OCT-1999;	99US-0162142P.	PR	16-JUN-1999;	99US-0139452P.
PR	29-OCT-1999;	99US-0162142P.	PR	16-JUN-1999;	99US-0139453P.
PR	29-OCT-1999;	99US-0162142P.	PR	17-JUN-1999;	99US-0139492P.
PR	29-OCT-1999;	99US-0162142P.	PR	18-JUN-1999;	99US-0139454P.
PR	29-OCT-1999;	99US-0162142P.	PR	18-JUN-1999;	99US-0139455P.
PR	29-OCT-1999;	99US-0162142P.	PR	18-JUN-1999;	99US-0139456P.
PR	29-OCT-1999;	99US-0162142P.	PR	18-JUN-1999;	99US-0139461P.
PR	29-OCT-1999;	99US-0162142P.	PR	18-JUN-1999;	99US-0139462P.
PR	29-OCT-1999;	99US-0162142P.	PR	18-JUN-1999;	99US-0139463P.
PR	29-OCT-1999;	99US-0162142P.	PR	18-JUN-1999;	99US-0139750P.
PR	29-OCT-1999;	99US-0162142P.	PR	18-JUN-1999;	99US-0139763P.
PR	29-OCT-1999;	99US-0162142P.	PR	21-JUN-1999;	99US-0139817P.
PR	29-OCT-1999;	99US-0162142P.	PR	22-JUN-1999;	99US-0139899P.
PR	29-OCT-1999;	99US-0162142P.	PR	23-JUN-1999;	99US-0140353P.
PR	29-OCT-1999;	99US-0162142P.	PR	23-JUN-1999;	99US-0140354P.
PR	29-OCT-1999;	99US-0162142P.	PR	24-JUN-1999;	99US-0140695P.
PR	29-OCT-1999;	99US-0162142P.	PR	28-JUN-1999;	99US-0140823P.
PR	29-OCT-1999;	99US-0162142P.	PR	29-JUN-1999;	99US-0140991P.
PR	29-OCT-1999;	99US-0162142P.	PR	30-JUN-1999;	99US-0141287P.
PR	29-OCT-1999;	99US-0162142P.	PR	01-JUL-1999;	99US-0141842P.
PR	29-OCT-1999;	99US-0162142P.	PR	01-JUL-1999;	99US-0142154P.
PR	29-OCT-1999;	99US-0162142P.	PR	02-JUL-1999;	99US-0142055P.
PR	29-OCT-1999;	99US-0162142P.	PR	06-JUL-1999;	99US-0142390P.
PR	29-OCT-1999;	99US-0162142P.	PR	08-JUL-1999;	99US-0142803P.
PR	29-OCT-1999;	99US-0162142P.	PR	09-JUL-1999;	99US-0142920P.
PR	29-OCT-1999;	99US-0162142P.	PR	12-JUL-1999;	99US-0142977P.
PR	29-OCT-1999;	99US-0162142P.	PR	13-JUL-1999;	99US-0143542P.
PR	29-OCT-1999;	99US-0162142P.	PR	14-JUL-1999;	99US-0143624P.
PR	29-OCT-1999;	99US-0162142P.	PR	15-JUL-1999;	99US-0144005P.
PR	29-OCT-1999;	99US-0162142P.	PR	16-JUL-1999;	99US-0144085P.
PR	29-OCT-1999;	99US-0162142P.	PR	16-JUL-1999;	99US-0144086P.
PR	29-OCT-1999;	99US-0162142P.	PR	19-JUL-1999;	99US-0144325P.
PR	29-OCT-1999;	99US-0162142P.	PR	19-JUL-1999;	99US-0144331P.
PR	29-OCT-1999;	99US-0162142P.	PR	19-JUL-1999;	99US-0144332P.
PR	29-OCT-1999;	99US-0162142P.	PR	19-JUL-1999;	99US-0144333P.

Query Match 60.3%; Score 41; DB 3; Length 252;

Best Local Similarity 46.2%; Pred. No. 21;

Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

OY 1 DWRPRTSYLLEI 13

Db | : | : | : | : |

77 DTHPHMSRYIQL 89

RESULT 11

AAG0655

ID AAG06655 standard; protein; 272 AA.

XX AC

XX AAG06655;

XX AC

DT 17-OCT-2000 (first entry)

XX DE

XX Arabidopsis thaliana protein fragment SEQ ID NO: 3506.

XX KW

KW Protein identification; signal transduction pathway; metabolic pathway;

KW hybridisation assay; genetic mapping; gene expression control; promoter;

KW termination sequence.

XX OS

XX Arabidopsis thaliana.

XX FN

XX EPI033405-A2.

XX XX

PD 06-SEP-2000.

XX XX

XX 25-FEB-2000; 2000EP-00301439.

XX XX

PR 25-FEB-1999;

PR 05-MAR-1999;

PR 05-MAR-1999;

PR 23-MAR-1999;

PR 25-MAR-1999;

PR 29-MAR-1999;

PR 01-APR-1999;

PR 06-APR-1999;

PR 08-APR-1999;

PR 99US-0121825P.

PR 99US-0123180P.

PR 99US-0123548P.

PR 99US-0125788P.

PR 99US-0126264P.

PR 99US-0126785P.

PR 99US-0127452P.

PR 99US-0128234P.

PR 99US-0128714P.



99US-0145088P  
99US-0145089P  
99US-0145087P  
99US-0145089P  
99US-0145132P  
99US-0145145P  
99US-0145214P  
99US-0145224P  
99US-0145276P  
99US-0145317P  
99US-0145318P  
99US-0145319P  
99US-0145391P  
99US-0146336P  
99US-0146388P  
99US-0146389P  
99US-0146388P  
99US-0147036P  
99US-0147204P  
99US-0147302P  
99US-0147304P  
99US-0147192P  
99US-0147260P  
99US-0147303P  
99US-0147416P  
99US-0147493P  
99US-0147535P  
99US-0147535P  
99US-0148171P  
99US-0148313P  
99US-0148341P  
99US-0148565P  
99US-0148684P  
99US-0149368P  
99US-0149175P  
99US-0149426P  
99US-0149722P  
99US-0149723P  
99US-0149928P  
99US-0149902P  
99US-0149930P  
99US-0149930P  
99US-0150684P  
99US-0151065P  
99US-0151065P  
99US-0151080P  
99US-0151033P  
99US-0151436P  
99US-0151930P  
99US-0152363P  
99US-0152363P  
99US-0153070P  
99US-0153759P  
99US-0154039P  
99US-0154039P  
99US-0154779P  
99US-0155139P  
99US-0155148P  
99US-0155645P  
99US-0155659P  
99US-0156591P  
99US-0156591P  
99US-0157753P  
99US-0157753P  
99US-0157866P  
99US-0158029P  
99US-0158029P  
99US-0158236P  
99US-0159293P  
99US-0159293P  
99US-0159294P  
99US-0159295P  
99US-0159329P  
99US-0159329P  
99US-0159331P  
99US-0159637P  
99US-0159638P  
99US-0159684P  
99US-0160741P



PT for homologous nucleic acids required for cellular proliferation to  
PT isolate candidate molecules for rational drug discovery programs.  
XX  
PS Claim 25; SEQ ID NO 73158; 1766pp; English.  
XX  
CC The invention relates to an isolated nucleic acid comprising any one of  
CC the 6213 antisense sequences given in the specification where expression  
CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
CC (1) a vector comprising a promoter operably linked to the nucleic acid  
CC encoding a polypeptide whose expression is inhibited by the antisense  
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
CC polypeptide or its fragment whose expression is inhibited by the  
CC antisense nucleic acid; (4) an antibody capable of specifically binding  
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
CC proliferation or the activity of a gene in an operon required for  
CC proliferation; (7) identifying a compound that influences the activity of  
CC the gene product or that has an activity against a biological pathway  
CC required for proliferation, or that inhibits cellular proliferation; (8)  
CC identifying a gene required for cellular proliferation or the biological  
CC pathway in which a proliferation-required gene or its gene product lies  
CC or a gene on which the test compound that inhibits proliferation of an  
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of  
CC strains; or (13) identifying the target of a compound that inhibits the  
CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
CC the target prokaryotic essential genes. Note: The sequence data for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from WIPO at  
XX ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 944 AA;  
  
Query Match 60.3%; Score 41; DB 6; Length 944;  
Best Local Similarity 70.0%; Pred. No. 92;  
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
  
QY 4 PYRTSRYLEI 13  
Db 115 PYRTAFYLEV 124  
  
RESULT 15  
ABU32202  
ID ABU32202 standard; protein; 961 AA.  
XX  
AC ABU32202;  
XX  
XX 19-JUN-2003 (first entry)  
DT  
XX Protein encoded by Prokaryotic essential gene #17729.  
DE  
XX Antisense; prokaryotic essential gene; cell proliferation; drug design.  
XX  
XX Klebsiella pneumoniae.  
OS  
XX WO200277183-A2.  
XX  
XX 03-OCT-2002.  
XX  
XX 21-MAR-2002; 2002WO-US009107.  
PF  
XX 21-MAR-2001; 2001US-00815242.  
PR  
XX 06-SEP-2001; 2001US-00948993.  
PR  
XX 25-OCT-2001; 2001US-0342923P.  
PR  
XX 08-FEB-2002; 2002US-00072851.  
PR  
XX 06-MAR-2002; 2002US-0362699P.

XX (ELIT-) ELITRA PHARM INC.  
XX Wang L, Zamudio C, Malone C, Hasselbeck R, Ohlsen KL, Zyskind JW;  
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
XX  
DR WPI; 2003-029926/02.  
DR N-PSDB; ACA36072.  
XX  
PT New antisense nucleic acids, useful for identifying proteins or screening  
PT for homologous nucleic acids required for cellular proliferation to  
PT isolate candidate molecules for rational drug discovery programs.  
XX  
PS Claim 25; SEQ ID NO 60126; 1766pp; English.  
XX  
CC The invention relates to an isolated nucleic acid comprising any one of  
CC the 6213 antisense sequences given in the specification where expression  
CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
CC (1) a vector comprising a promoter operably linked to the nucleic acid  
CC encoding a polypeptide whose expression is inhibited by the antisense  
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
CC polypeptide or its fragment whose expression is inhibited by the  
CC antisense nucleic acid; (4) an antibody capable of specifically binding  
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
CC proliferation or the activity of a gene in an operon required for  
CC proliferation; (7) identifying a compound that influences the activity of  
CC the gene product or that has an activity against a biological pathway  
CC required for proliferation, or that inhibits cellular proliferation; (8)  
CC identifying a gene required for cellular proliferation or the biological  
CC pathway in which a proliferation-required gene or its gene product lies  
CC or a gene on which the test compound that inhibits proliferation of an  
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of  
CC strains; or (13) identifying the target of a compound that inhibits the  
CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
CC the target prokaryotic essential genes. Note: The sequence data for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from WIPO at  
XX ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 961 AA;  
  
Query Match 60.3%; Score 41; DB 6; Length 961;  
Best Local Similarity 70.0%; Pred. No. 94;  
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
  
QY 4 PYRTSRYLEI 13  
Db 124 PYRTAFYLEV 133  
  
Search completed: October 5, 2004, 07:59:38  
Job time : 9.89043 secs

This Page Blank (uspio)



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2004, 07:36:16 ; Search time 2.10648 Seconds  
(without alignments)  
593.639 Million cell updates/sec

Title: US-09-805-290A-11

Perfect score: 68

Sequence: 1 QVRVRFSSDYNTY 13

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 78:\*

1: p1r1:\*

2: p1r2:\*

3: p1r3:\*

4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	42	61.8	1225	2 T39255	probable C2 domain
2	40	58.8	152	2 AF2416	hypothetical prote
3	39	57.4	407	2 AB1205	molycoprotein bios
4	39	57.4	470	2 D69394	phosphoribosylamin
5	39	57.4	954	2 I61714	co-repressor prote
6	38	55.9	317	2 G90878	probable transcrip
7	38	55.9	317	2 C85529	probable AraC-like
8	38	55.9	334	2 T29672	hypothetical prote
9	38	55.9	452	2 A12071	hypothetical prote
10	38	55.9	2237	2 T21087	hypothetical prote
11	38	55.9	2245	2 T21086	hypothetical prote
12	37	54.4	72	2 I51369	Sry-related sequen
13	37	54.4	92	2 D96947	hypothetical prote
14	37	54.4	232	2 G70355	phosphoribosylanth
15	37	54.4	281	2 G71095	hypothetical prote
16	37	54.4	479	2 A38165	lactaldehyde dehyd
17	37	54.4	479	2 E90881	aldehyde dehydroge
18	37	54.4	479	2 D85737	aldehyde dehydroge
19	37	54.4	508	2 G83977	L-2,4-diaminobuty
20	37	54.4	684	2 I39595	phospholipase C -
21	36.5	53.7	879	2 A47704	endoglucanase I (B
22	36.5	53.7	915	2 A43802	cellulase (EC 3.2.
23	36.5	53.7	1039	2 S02711	cellulase (EC 3.2.
24	36.5	53.7	1331	2 A48954	mannan endo-1,4-be
25	36.5	53.7	1711	2 T31337	1,4-beta-glucanase
26	36	52.9	71	2 I50541	Sry-related sequen
27	36	52.9	119	2 H95178	general stress pro
28	36	52.9	128	2 G98045	conserved hypotat
29	36	52.9	131	2 E69488	SSU ribosomal prot

30 36 52.9 131 2 A05144 beta-lactoglobulin  
31 36 52.9 174 2 B81932 outer membrane pro  
32 36 52.9 174 2 G81174 outer membrane pro  
33 36 52.9 293 2 A97396 hypothetical prote  
34 36 52.9 293 2 AB2614 hypothetical prote  
35 36 52.9 333 2 T36397 hypothetical prote  
36 36 52.9 333 2 F81955 lactaldehyde dehyd  
37 36 52.9 480 2 A81023 aldehyde dehydroge  
38 36 52.9 511 2 A06636 periplasmic glucan  
39 36 52.9 547 2 C87992 protein W09G3.4 [1  
40 36 52.9 875 2 T17382 vrlC protein - Dic  
41 36 52.9 1626 2 T26318 hypothetical prote  
42 36 52.9 1718 1 JQ1734 genome polyprotein  
43 36 52.9 2139 2 S46404 vitellogenin - yel  
44 35 51.5 29 1 GCEN glucagon - elephan  
45 35 51.5 36 1 GCPI glucagon-36 - spot

#### ALIGNMENTS

##### RESULT 1

T39255  
probable C2 domain family protein - fission yeast (Schizosaccharomyces pombe)  
C:Species: Schizosaccharomyces pombe  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 03-Dec-1999  
C:Accession: T39255  
R:Stevens, K.; Church, C.; McDougall, R.; Rajandream, M.A.; Barrell, B.G.  
submitted to the EMBL Data Library, August 1999  
A:Reference number: Z21838  
A:Accession: T39255  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-1225 <STE>  
A:Cross-references: EMBL:ALJ09734; PIDN:CABS2146.1; GSPDB:GN00066; SPDB:SPAPYUK71.03c  
A:Experimental source: strain 972h-; cosmid pYUK71  
C:Genetics:  
A:Gene: SPDB:SPAPYUK71.03c  
A:Map position: 1

Query Match 61.8%; Score 42; DB 2; Length 1225;  
Best Local Similarity 53.8%; Pred. No. 24;  
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 QVRVRFSSDYNTY 13  
Db 196 RVKVFDDYTRY 208

##### RESULT 2

AF2416  
hypothetical protein all4886 [imported] - Nostoc sp. (strain PCC 7120)  
C:Species: Nostoc sp. PCC 7120  
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
C:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 09-Dec-2002  
C:Accession: AF2416  
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iziguchi, N.; Nakazaki, N.; Shimpou, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tasuda, M.; Tabata, S.  
DNA Res. 8, 205-213, 2001  
A:Title: Complete Genomic  
A:Reference number: AB1807; MUID:21595285; PMID:11759840  
A:Accession: AF2416  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-152 <KUR>  
A:Cross-references: GB:BA000019; PIDN:BAB76585.1; PID:gl7134023; GSPDB:GN00179  
A:Experimental source: strain PCC 7120  
C:Genetics:  
A:Gene: all4886

Query Match 58.8%; Score 40; DB 2; Length 152;  
Best Local Similarity 54.5%; Pred. No. 6.1;  
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;



Query Match 55.9%; Score 38; DB 2; Length 317;  
 Best Local Similarity 58.3%; Pred. No. 31;  
 Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 VRVRFSSDYTN 13  
 |||:|||||:  
 Db 170 VRQHFQSDIANH 181

RESULT 8  
 T29672  
 hypothetical protein C50E3.4 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 04-Mar-2000  
 C:Accession: T29672  
 R:Geisel, C.; Bradshaw, H.  
 submitted to the EMBL Data Library, July 1996  
 A:Description: The sequence of C. elegans cosmid C50E3.  
 A:Reference number: Z20662  
 A:Accession: T29672  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-334 <GEI>  
 A:Cross-references: EMBL:U64848; PIDN:AA04881.1; GSPDB:GN00023; CESP:C50E3.4  
 A:Experimental source: strain Bristol N2; clone C50E3  
 C:Genetics:  
 A:Gene: CESP:C50E3.4  
 A:Map position: 5  
 A:Introns: 15/3; 53/1; 110/3; 140/2; 276/2  
 C:Superfamily: Caenorhabditis elegans hypothetical protein C50E3.4

Query Match 55.9%; Score 38; DB 2; Length 334;  
 Best Local Similarity 54.5%; Pred. No. 32;  
 Matches 6; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

QY 2 VRVRFSSDYTN 12  
 |||:|||||:  
 Db 13 VRVKFASFSN 23

RESULT 9  
 AI2071  
 hypothetical protein all2127 [imported] - Nostoc sp. (strain PCC 7120)  
 C:Species: Nostoc sp. PCC 7120  
 A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
 C>Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 09-Dec-2002  
 C:Accession: AI2071  
 R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S.  
 DNA Res. 8, 205-213, 2001  
 A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena  
 A:Reference number: AB1807; MUID:21595285; PMID:11759840  
 A:Accession: AI2071  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-452 <KUR>  
 A:Cross-references: GB:BA000019; PIDN:BA073826.1; PID:GL7131218; GSPDB:GN00179  
 A:Experimental source: strain PCC 7120  
 C:Genetics:  
 A:Gene: all2127

Query Match 55.9%; Score 38; DB 2; Length 452;  
 Best Local Similarity 46.2%; Pred. No. 45;  
 Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 QVRVRFSSDYTN 13  
 |||:|||||:  
 Db 246 QINIRCNVDYQNY 258

RESULT 10  
 T21087  
 hypothetical protein F18C12.2b - Caenorhabditis elegans

C:Species: Caenorhabditis elegans  
 C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
 C:Accession: T21087  
 R:Harris, B.  
 submitted to the EMBL Data Library, June 1996  
 A:Reference number: Z19371  
 A:Accession: T21087  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-2237 <WIL>  
 A:Cross-references: EMBL:Z75536; PIDN:CAA99832.1; GSPDB:GN00019; CESP:F18C12.2b  
 C:Genetics:  
 A:Gene: CESP:F18C12.2b  
 A:Map position: 1  
 A:Introns: 48/3; 115/3; 181/3; 282/3; 369/1; 524/3; 661/3; 760/3; 839/2; 1065/2; 1278/3;

Query Match 55.9%; Score 38; DB 2; Length 2237;  
 Best Local Similarity 87.5%; Pred. No. 2.4e+02;  
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 4 VRFSDDYT 11  
 :|||||:  
 Db 86 MRFSSDYT 93

RESULT 11  
 T21086  
 hypothetical protein F18C12.2a - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
 C:Accession: T21086  
 R:Harris, B.  
 submitted to the EMBL Data Library, June 1996  
 A:Reference number: Z19371  
 A:Accession: T21086  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-2245 <WIL>  
 A:Cross-references: EMBL:Z75536; PIDN:CAA99831.1; GSPDB:GN00019; CESP:F18C12.2a  
 A:Experimental source: clone F18C12  
 C:Genetics:  
 A:Gene: CESP:F18C12.2a  
 A:Map position: 1  
 A:Introns: 48/3; 115/3; 181/3; 282/3; 369/1; 524/3; 661/3; 760/3; 839/2; 1065/2; 1278/3;

Query Match 55.9%; Score 38; DB 2; Length 2245;  
 Best Local Similarity 87.5%; Pred. No. 2.4e+02;  
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 4 VRFSDDYT 11  
 :|||||:  
 Db 86 MRFSSDYT 93

RESULT 12  
 I51369  
 Sry-related sequence - Tarentola mauritanica (fragment)  
 C:Species: Tarentola mauritanica  
 C>Date: 04-Sep-1997 #sequence\_revision 07-Nov-1997 #text\_change 24-Sep-1999  
 C:Accession: I51369  
 R:Picardo, A.M.; Mueller, U.; Harry, J.L.; Uwanogho, D.; Sharpe, P.T.  
 PCR Methods Appl. 2, 218-222, 1992  
 A:Title: PCR amplification of SRY-related gene sequences reveals evolutionary conservatism  
 A:Reference number: I50019; MUID:93184703; PMID:8443573  
 A:Accession: I51369  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-72 <PIC>  
 A:Cross-references: GB:M86337; NID:G213877; PIDN:AAA49621.1; PID:G213878  
 C:Genetics:  
 A:Gene: MG42  
 C:Superfamily: unassigned HMG box proteins; HMG box homology

F1:72/Domain: HMG box homology (fragment) <HMG>

Query Match 54.4%; Score 37; DB 2; Length 72;  
Best Local Similarity 46.2%; Pred. No. 9.9;  
Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 QVRVRFSSDYTNV 13  
::||: :|||  
Db 56 RLVRKHMADYFNY 68

#### RESULT 13

D96947 Hypothetical protein CAC0387 [imported] - Clostridium acetobutylicum

C:Species: Clostridium acetobutylicum  
C:Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 14-Sep-2001  
C:Accession: D96947  
R:Noaling, J.; Breton, G.; Onelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee, J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.  
J. Bacteriol. 193, 4823-4838, 2001  
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Cld  
A:Reference number: A96900; MUID:21359325; PMID:21359325  
A:Accession: D96947  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-92 <KUR>  
A:Cross-references: GB:AE001437; PIDN:AAK78367.1; PID:G15023237; GSPDB:GN00168  
A:Experimental source: Clostridium acetobutylicum ATCC824  
C:Genetics:  
A:Gene: CAC0387

Query Match 54.4%; Score 37; DB 2; Length 92;  
Best Local Similarity 46.2%; Pred. No. 13;  
Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 QVRVRFSSDYTNV 13  
::||: :|||  
Db 16 KIEYSFSTDYFNY 28

#### RESULT 14

S70355 phosphoribosylanthranilate isomerase (EC 5.3.1.24) - Lipomyces starkeyi

C:Species: Lipomyces starkeyi  
C:Date: 06-Dec-1996 #sequence\_revision 25-Apr-1997 #text\_change 20-Jun-2000  
C:Accession: S70355  
R:Bignelli, G.R.; Bruce, I.J.; Evans, I.H.  
Curr. Genet. 30, 83-88, 1996  
A:Title: Electrophoretic karyotype of the amyolytic yeast Lipomyces starkeyi and cloning  
A:Reference number: S70355; MUID:96269934; PMID:8662214  
A:Accession: S70355  
A:Molecule type: DNA  
A:Residues: 1-232 <BIG>  
A:Cross-references: EMBL:Z68292; NID:G1134847; PIDN:CAA92584.1; PID:G1161576  
A:Note: the authors translated the codon GCC for residue 211 as Glu  
C:Genetics:  
A:Gene: TRP1

C:Superfamily: phosphoribosylanthranilate isomerase; trpF homology  
C:Keywords: intramolecular oxidoreductase; isomerase; tryptophan biosynthesis  
F.14-228/Domain: trpF homology <TRF>

Query Match 54.4%; Score 37; DB 2; Length 232;  
Best Local Similarity 58.3%; Pred. No. 34;  
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 2 VVRVRFSSDYTNV 13  
||| :||:|  
Db 80 VRGTFSDYTSW 91

#### RESULT 15

G71095 hypothetical protein PH1028 - Pyrococcus horikoshii

C:Species: Pyrococcus horikoshii  
C:Date: 14-Aug-1998 #sequence\_revision 14-Aug-1998 #text\_change 05-Nov-1999  
C:Accession: G71095  
R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Sekine, M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi, DNA Res. 5, 55-76, 1998  
A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic ar  
A:Reference number: A71000; MUID:98344137; PMID:9679194  
A:Accession: G71095  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-281 <KAW>  
A:Cross-references: GB:AP000004; NID:G3236131; PIDN:BAA30125.1; PID:d1031068; PID:G325744  
A:Experimental source: strain OT3  
A:Note: this accession replaces an interim accession for a sequence replaced by GenBank  
C:Genetics:  
A:Gene: PH1028

Query Match 54.4%; Score 37; DB 2; Length 281;  
Best Local Similarity 46.2%; Pred. No. 41;  
Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 QVRVRFSSDYTNV 13  
::||: :|||  
Db 33 KIRVRVSKDYVEF 45

Search completed: October 5, 2004, 08:16:54  
Job time : 6.10648 secs

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2004, 07:29:35 ; Search time 1.16358 Seconds  
(without alignments)  
581.749 Million cell updates/sec

Title: US-09-805-290A-11  
Perfect score: 68  
Sequence: 1 QVRVRFSSDYTNV 13

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	42	61.8	1225	1 YKH3 SCHPO	Q9UT00 schizosacch
2	39	57.4	470	1 PUR2 ARCFU	Q29108 archaeoglob
3	39	57.4	954	1 SN3B MOUSE	Q62141 mus musculus
4	39	57.4	1130	1 SN3B HUMAN	O75182 homo sapien
5	37	54.4	72	1 MG42 TARMA	P40651 tarentola m
6	37	54.4	232	1 TRPE_LIPST	Q01128 lipomyces s
7	37	54.4	478	1 ALDA_ECOLI	P25553 escherichia
8	37	54.4	1033	1 ATHL_CAVPO	Q64392 cavia porce
9	37	54.4	1169	1 C8BA_BACUK	Q45705 bacillus th
10	36.5	53.7	879	1 GUNI_CLOTM	Q02934 clostridium
11	36.5	53.7	1039	1 GUNB_CALSA	P10474 c endocluca
12	36.5	53.7	1331	1 MANE_CALSA	P22533 caldocellum
13	36	52.9	71	1 LG27_EUBMA	P40654 eubieparis
14	36	52.9	131	1 RS8_ARCFU	Q28369 archaeoglob
15	36	52.9	395	1 RT31 HUMAN	Q92665 homo sapien
16	36	52.9	511	1 OPGG_SALTU	O8xiff salmonella
17	36	52.9	691	1 SYGB_BUCAP	O8ka07 buchneza ap
18	36	52.9	1718	1 RRPO_SHVX	Q04575 shallow vir
19	36	52.9	2148	1 VITI_AEDAE	Q16927 aedes segyp
20	35	51.5	29	1 GLUC_CALMI	P13189 callorhynch
21	35	51.5	36	1 GLUC_HYDCO	F09682 hydrolegus
22	35	51.5	178	1 YE57_HAEIN	F05720 haemophilus
23	35	51.5	181	1 E320_ADE1A	P35767 human adeno
24	35	51.5	181	1 E320_ADE1P	P35768 human adeno
25	35	51.5	181	1 E320_ADE3S	P15317 human adeno
26	35	51.5	294	1 LYCM_STRGL	P25310 streptomyc
27	35	51.5	343	1 ARGC_STAMM	Q99X37 staphylococ
28	35	51.5	602	1 GP63_LEIMA	P08148 leishmania
29	35	51.5	1694	1 CLH_DICDI	P25870 dictyostell
30	35	51.5	3564	1 CSM1_MOUSE	Q92313 mus musculu
31	35	51.5	3565	1 CSM1_HUMAN	Q96P37 homo sapien
32	34	50.0	72	1 MG42 TARMA	P40653 tarentola m
33	34	50.0	107	1 Y339_METJA	Q57785 methanococc

## ALIGNMENTS

## RESULT 1

YKH3	SCHPO				
ID	YKH3_SCHPO	STANDARD;	PR	1225	AA.
AC	Q9UT00;				
DT	28-FEB-2003	(Rel. 41, Created)			
DT	28-FEB-2003	(Rel. 41, Last sequence update)			
DT	28-FEB-2003	(Rel. 41, Last annotation update)			
DE	Hypothetical protein PYUK71.03c in chromosome I.				
GN	SPAPYUK71.03C.				
OS	Schizosaccharomyces pombe (Fission yeast).				
OC	Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;				
OC	Schizosaccharomycetales; Schizosaccharomycetaceae;				
OC	Schizosaccharomycetes.				
OX	NCBI_TaxID=4896;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=972;				
RC	MEDLINE=21848401; PubMed=11859360;				
RA	Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,				
RA	Sgourou J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,				
RA	Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,				
RA	Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,				
RA	Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,				
RA	Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,				
RA	James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,				
RA	Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,				
RA	Oliver K., O'Neil S., Pearson D., Quail M.A., Rabinowitsch E.,				
RA	Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,				
RA	Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,				
RA	Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,				
RA	Woodward J., Volkart G., Aert R., Robben J., Grymonprez B.,				
RA	Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,				
RA	Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,				
RA	Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,				
RA	Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,				
RA	Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,				
RA	Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,				
RA	Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,				
RA	Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,				
RA	Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,				
RA	Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,				
RA	Shpakovski G.V., Useery D., Barrell B.G., Nurse P.;				
RT	"The genome sequence of Schizosaccharomyces pombe.";				
RT	Nature 415:871-880(2002).				
CC	-!- SIMILARITY: Contains 3 C2 domains.				
CC	This SWISS-PROT entry is copyright. It is produced through a collaborati				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation				
CC	the European Bioinformatics Institute. There are no restrictions on i				
CC	use by non-profit institutions as long as its content is in no				
CC	modified and this statement is not removed. Usage by and for commerci				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announc				
CC	or send an email to license@isb-sib.ch).				
CC					
CC					
CC					
CC					
CC					
CC					
CC					
CC					
CC					
CC					
CC					
CC					
CC					
CC					
CC					
CC					
CC					
CC					
CC					
CC					
CC					
CC					
CC					
CC					
CC					
CC					
CC					
CC					
CC					
CC					
CC					
CC					
CC					
CC					
CC					
CC					
CC					
CC					
CC					
CC					
CC					
CC					
CC					
CC					
CC					
CC					
CC					
CC					
CC					
CC					
CC					
CC					
CC					
CC					
CC					
CC					
CC					
CC					
CC					
CC					
CC					
CC					
CC					
CC					
CC					
CC					
CC					
CC					
CC					
CC					
CC					
CC					
CC					
CC					
CC					
CC					
CC					
CC					
CC					
CC					
CC					
CC					
CC					
CC					
CC					
CC					
CC					
CC					
CC					
CC					
CC					
CC					
CC					
CC					
CC					
CC					
CC					
CC					
CC					
CC					
CC					
CC					
CC					
CC					
CC					
CC					
CC					
CC					
CC					
CC					
CC					
CC					
CC					

```

DR HSP: P21707; 1BYN.
DR GensDB_Spome; SPAPYUK71.03c; -.
DR InterPro; IPR000008; C2.
DR InterPro; IPR008973; C2_CalB.
DR Pfam; PF00168; C2; 4.
DR PRINTS; PR00360; C2DOMAIN.
DR SMART; SM00239; C2; 4.
DR PROSITE; PS00499; C2_DOMAIN 1; 1.
DR PROSITE; PS50004; C2_DOMAIN 2; 3.
KW Hypothetical protein; Transmembrane; Repeat.
FT TRANSMEM 167 187 POTENTIAL.
FT DOMAIN 414 518 C2 DOMAIN 1.
FT DOMAIN 691 787 C2 DOMAIN 2.
FT DOMAIN 1024 1121 C2 DOMAIN 3.
SQ SEQUENCE 1225 AA; 135781 MW; 7041FCRA5A6D7B96 CRC64;

Query Match 61.8%; Score 42; DB 1; Length 1225;
Best Local Similarity 53.8%; Pred. NO. 11;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 QVRVFSSDYTNV 13
DQ 196 RVKNFKDDYTRY 208

RESULT 2
ID PUR2_ARCFU STANDARD; PRT; 470 AA.
AC Q29108.
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Phosphoribosylamine-glycine ligase (EC 6.3.4.13) (GARS) (Glycinamide
DE ribonucleotide synthetase) (Phosphoribosylglycinamide synthetase).
GN PURD OR AF1157.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
OC Archaeoglobaceae; Archaeoglobus.
OX NCBI_TaxID=2234;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VC-15 / DSM 4304 / ATCC 49558;
RX MEDLINE=98049343; PubMed=9389475;
RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Ketchum K.A., Dodson R.J., Gwinn M., Hickley E.K., Peterson J.D.,
RA Richardson D.L., Kerlavage A.R., Graham D.E., Kyrpides N.C.,
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Uterback T.,
RA Cotton M.D., Spriggs T., Artlich P., Kaine B.P., Sykes S.M.,
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RA "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon Archaeoglobus fulgidus.";
RL Nature 390:364-370(1997).
CC -!- CATALYTIC ACTIVITY: ATP + 5-phospho-D-ribosylamine + glycine = ADP
CC + phosphate + N(1)-(5-phospho-D-ribose)glycinamide.
CC -!- PATHWAY: De novo purine biosynthesis; second step.
CC -!- SIMILARITY: Belongs to the GARS family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institutes of Bioinformatics and the EMBL outstation.
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE001024; AAB90089.1; -.
CC PIR; D69394; D69394.
CC HSP; F15640; IGSO.

```

---

```

DR TIGR; AF1157; -.
DR HAMAP; MF_00138; atypical; 1.
DR InterPro; IPR000115; Gars.
DR Pfam; PF01071; GARS; 1.
DR Pfam; PF02842; GARS_B; 1.
DR Pfam; PF02843; GARS_C; 1.
DR Pfam; PF02844; GARS_N; 1.
DR TIGRFAMS; TIGR00877; purD; 1.
DR PROSITE; PS00184; GARS; FALSE NEG.
KW Purine biosynthesis; Ligase; Complete proteome.
FT DOMAIN 230 260 INSERT.
SQ SEQUENCE 470 AA; 51750 MW; C3B90CED22DCD353 CRC64;

Query Match 57.4%; Score 39; DB 1; Length 470;
Best Local Similarity 70.0%; Pred. NO. 14;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 QVRVFSSDY 10
DQ 359 EVEMRPFSSDY 368

RESULT 3
ID SN3B_MOUSE STANDARD; PRT; 954 AA.
AC Q62141; OS4976; Q8VCB8; Q8VDZ5;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Paired amphipathic helix protein Sin3b.
GN SIN3B.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1), FUNCTION, SUBUNIT, AND SUBCELLULAR
RP LOCATION.
RC TISSUE=Kidney;
RX MEDLINE=95196269; PubMed=7889570;
RA Ayer D.E., Lawrence O.A., Bisenman R.N.;
RT "Mad-Max transcriptional repression is mediated by ternary complex
RT formation with mammalian homologs of yeast repressor Sin3.";
RL Cell 80:767-776(1995).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 2), FUNCTION, AND INTERACTION WITH FOXK1.
RC TISSUE=Heart;
RX MEDLINE=20088666; PubMed=10620510;
RA Yang Q., Kong Y., Rothermel B., Garry D.J., Bassel-Duby R.,
RA Williams R.S.;
RT "The winged-helix/forkhead protein myocyte nuclear factor beta (MNF-
RT beta) forms a co-repressor complex with mammalian Sin3B.";
RL Biochem. J. 345:335-343(2000).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORMS 2 AND 3).
RC TISSUE=Mammary gland, and Salivary gland;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan B., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.A., Loquellano N.A., Peters G.J., Abramson R.D., Maltahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalski U., Smailus D.E.,

```

RA Schmerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 RL and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [4]  
 RP STRUCTURE BY NMR OF 148-232 IN COMPLEX WITH MAD1.  
 RX MEDLINE=20553704; PubMed=11101889;  
 RA Spronk C.A.E., Tessari M., Kaa A.M., Jansen J.F.A., Vermeulen M.,  
 RA Stunnenberg H.G., Vuister G.W.;  
 RT "The Mad1-Sin3b interaction involves a novel helical fold.";  
 RL Nat. Struct. Biol. 7:1100-1104(2000).  
 CC -!- FUNCTION: Acts as a transcriptional repressor. Interacts with MXII  
 CC to repress MYC responsive genes and antagonize MYC oncogenic  
 CC activities. Interacts with MAD-MAX heterodimers by binding to MAD.  
 CC The heterodimer then represses transcription by tethering SIN3B to  
 CC DNA. Also forms a complex with FOXK1 which represses  
 CC transcription.  
 CC -!- SUBUNIT: Interacts with FOXK1/MNF, MXII and MAD.  
 CC -!- SUBCELLULAR LOCATION: Nuclear.  
 CC -!- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=3;  
 CC Name=1;  
 CC IsoId=Q62141-1; Sequence=Displayed;  
 CC Name=2;  
 CC IsoId=Q62141-2; Sequence=VSP\_008225, VSP\_008226;  
 CC Name=3;  
 CC IsoId=Q62141-3; Sequence=VSP\_008227, VSP\_008228;  
 CC Note=No experimental confirmation available;  
 CC -!- SIMILARITY: Contains 3 PAH (paired amphipathic helix) repeats.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; L38622; AAA69774.1; -;  
 DR EMBL; AF038548; AAC04821.1; -;  
 DR EMBL; BC020049; AAH20049.1; -;  
 DR EMBL; BC021160; AAH21160.1; -;  
 DR PIR; I61714; I61714.  
 DR PDB; 1B91; 07-DEC-00.  
 DR TRASNFPAC; T02398; -;  
 DR MGD; MGII:107158; Sin3b.  
 DR GO; GO:0005634; C:nucleus; IDA.  
 DR GO; GO:0005515; F:protein binding; IPI.  
 DR GO; GO:0003714; F:transcription co-repressor activity; IDA.  
 DR GO; GO:0045892; P:negative regulation of transcription, DNA-d. . . ; IDA.  
 DR InterPro; IPR003822; PAH.  
 DR Pfam; PF02671; PAH; 3.  
 KW Transcription regulation; Repressor; Repeat; Nuclear protein;  
 KW Alternative splicing; 3D-structure.  
 FT REPEAT 52 98  
 FT REPEAT 172 228 PAH 1.  
 FT REPEAT 312 358 PAH 3.  
 FT VARSPPLIC 275 293  
 FT KKKLGGTKDLSTAAVGVKY -> VGLQLKCAVVVFGYCTAE  
 FT E (in isoform 2).  
 FT /FtId=VSP\_008225.  
 FT VARSPPLIC 294 954  
 FT Missing (in isoform 2).  
 FT /FtId=VSP\_008226.  
 FT VARSPPLIC 275 302  
 FT KKKLGGTKDLSTAAVGVGYTLCDFSF -> VLVHVVVLP  
 FT AKRGSGVEAQAAGEPEARA (in isoform 3).  
 FT /FtId=VSP\_008227.  
 FT VARSPPLIC 303 954  
 FT Missing (in isoform 3).  
 FT /FtId=VSP\_008228.  
 FT CONFLICT 230 230 A -> G (IN REF. 2).  
 FT CONFLICT 233 233 S -> P (IN REF. 2).  
 SQ SEQUENCE 954 AA; 109393 MW; 31F4BE14523EA213 CRC64;  
 Query Match 57.4%; Score 39; DB 1; Length 954;  
 Best Local Similarity 53.8%; Pred. No. 29;

Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;  
 QX 1 QVRFSSDYTNV 13  
 DB 43 QVIRFGSDPATY 55  
 RESULT 4  
 SN3B HUMAN  
 ID SN3B HUMAN STANDARD; PRT; 1130 AA.  
 AC O75132; Q8TB34; Q9BSC8;  
 DT 15-MAR-2004 (Rel. 43, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Paired amphipathic helix protein Sin3b.  
 GN SIN3B OR KIAA0700.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=98403880; PubMed=9734811;  
 RA Ishikawa K.-i., Nagase T., Suyama M., Miyajima N., Tanaka A.,  
 RA Kotani H., Nomura N., Ohara O.;  
 RT "Prediction of the coding sequences of unidentified human genes. X.  
 RT The complete sequences of 100 new cDNA clones from brain which can  
 RT code for large proteins in vitro.";  
 RL DNA Res. 5:169-176(1998).  
 RN [2]  
 RP SEQUENCE OF 2-353 AND 711-1130 FROM N.A.  
 RC TISSUE=Lung, and Skin;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldi M.F., Casavant T.L., Schetz T.E.,  
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallus D.E.,  
 RA Schmerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 CC -!- FUNCTION: Acts as a transcriptional repressor. Interacts with MXII  
 CC to repress MYC responsive genes and antagonize MYC oncogenic  
 CC activities. Interacts with MAD-MAX heterodimers by binding to MAD.  
 CC The heterodimer then represses transcription by tethering SIN3B to  
 CC DNA. Also forms a complex with FOXK1 which represses transcription  
 CC (By similarity).  
 CC -!- SUBUNIT: Interacts with FOXK1/MNF, MXII and MAD (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).  
 CC -!- SIMILARITY: Contains 3 PAH (paired amphipathic helix) repeats.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; AB014600; BAA31675.1; ALT\_INIT.





QY 1 QVRFVFSDDYNTY 13  
Db: 101 EVEVFTADYIDY 113

RESULT 8  
ATHL\_CAVPO STANDARD; PRT; 1033 AA.  
ID AC O64392;  
DT 15-JUL-1999 (Rel. 38, Created)  
DT 15-JUL-1999 (Rel. 38, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Potassium-transporting ATPase alpha chain 2 (EC 3.6.3.10) (Proton pump) (Non-gastric H<sup>+</sup>/K<sup>+</sup> ATPase alpha subunit).  
DE DE pump)  
GN ATP12A OR ATP1A1.  
OS Cavia porcellus (Guinea pig).  
CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
CC Mammalia; Eutheria; Rodentia; Hystricognathi; Caviidae; Cavia.  
OX NCBI\_TaxID=10141;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA STRAIN=Harley;  
RC Watanabe T., Sato M., Kaneko K., Suzuki T., Yoshida T., Suzuki Y.;  
RT "Isolation and characterization of cDNA encoding the putative guinea pig distal colon H<sup>+</sup>, K<sup>+</sup> ATPase alpha subunit.";  
RT Submitted (OCT-1993) to the EMBL/GenBank/DDJ databases.  
RL [2]  
RP TISSUE SPECIFICITY.  
RX MEDLINE=990687400; PubMed=98723395;  
RA Pestov N.B., Romanova L.G., Korzenko T.V., Egorov M.V., Kostina M.B.,  
RT Sverdlov V.B., Askari A., Shakipatov M.I., Modyanov N.N.;  
RT "Ouabain-sensitive H,K-ATPase: tissue-specific expression of the mammalian genes encoding the catalytic alpha subunit.";  
RT FEBS Lett. 440:320-324(1998).  
CC -I- FUNCTION: Catalyzes the hydrolysis of ATP coupled with the exchange of H(+) and K(+) ions across the plasma membrane.  
CC Responsible for potassium absorption in various tissues.  
CC -I- CATALYTIC ACTIVITY: ATP + H(2O) + H(+) (in) + K(+) (Out) = ADP + phosphate + H(+) (Out) + K(+) (in).  
CC -I- SUBUNIT: Composed of two subunits: alpha (catalytic) and beta.  
CC -I- SUBCELLULAR LOCATION: Integral membrane protein.  
CC -I- TISSUE SPECIFICITY: Found in skin, kidney and distal colon.  
CC -I- SIMILARITY: Belongs to the cation transport ATPases family (P-type ATPases). Subfamily IIC.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; D21854; BAA04880.1; --  
DR InterPro; IPR001757; ATPase\_E1-E2.  
DR InterPro; IPR006069; Cation\_ATPase.  
DR InterPro; IPR006068; Cation\_ATPase\_C.  
DR InterPro; IPR004014; Cation\_ATPase\_N.  
DR InterPro; IPR008250; E1-E2\_ATPase\_Reg.  
DR InterPro; IPR005834; HydroLase.  
DR InterPro; IPR005775; Na/K\_ATPase\_alph.  
DR Pfam; PF00689; Cation\_ATPase\_C; 1.  
DR Pfam; PF00690; Cation\_ATPase\_N; 1.  
DR Pfam; PF00122; E1-E2\_ATPase; 1.  
DR Pfam; PF00702; HydroLase; 1.  
DR PRINTS; PR00119; CATATPASE.  
DR PRINTS; PR00121; NAKATPASE.  
DR TIGRFAMs; TIGR01106; ATPase-IIC\_X-K; 1.  
DR TIGRFAMs; TIGR01494; ATPase\_P-type; 5.  
DR PROSITE; PS00154; ATPASE\_E1-E2; 1.  
DR HydroLase; Potassium transport; Hydrogen ion transport; Transmembrane; Phosphorylation; ATP-binding; Magnesium; Metal-binding.  
FT DOMAIN 1 96 CYTOPLASMIC (POTENTIAL).  
FT

DR PROSITE; PS00154; ATPASE\_E1\_E2; 1.  
KW Hydrolyase; Potassium transport; Hydrogen ion transport; Transmembrane  
KW Phosphorylation; ATP-binding; Magnesium; Metal-binding.  
FT DOMAIN 1 96 CYTOPLASMIC (POTENTIAL).  
96

```

FT TRANSMEM 97 117 POTENTIAL.
FT DOMAIN 118 140 LUMENAL (POTENTIAL).
FT TRANSMEM 141 161 POTENTIAL.
FT DOMAIN 162 297 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 298 317 POTENTIAL.
FT DOMAIN 318 329 LUMENAL (POTENTIAL).
FT TRANSMEM 330 347 POTENTIAL.
FT DOMAIN 348 781 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 782 801 POTENTIAL.
FT DOMAIN 802 811 LUMENAL (POTENTIAL).
FT TRANSMEM 812 832 POTENTIAL.
FT DOMAIN 833 852 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 853 875 POTENTIAL.
FT DOMAIN 876 927 LUMENAL (POTENTIAL).
FT TRANSMEM 928 947 POTENTIAL.
FT DOMAIN 948 961 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 962 980 POTENTIAL.
FT DOMAIN 981 995 LUMENAL (POTENTIAL).
FT TRANSMEM 996 1016 POTENTIAL.
FT DOMAIN 1017 1033 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 1034 1052 PHOSPHORYLATION (BY SIMILARITY).
FT MOD_RES 1053 385 PHOSPHORYLATION (BY PKA) (BY SIMILARITY).
FT MOD_RES 386 952 PHOSPHORYLATION (BY PKA) (BY SIMILARITY).
FT METAL 726 726 MAGNESIUM (BY SIMILARITY).
FT METAL 730 730 MAGNESIUM (BY SIMILARITY).
SQ SEQUENCE 1033 AA; 114557 MW; C4EBE72D83C1F40C CRC64;

Query Match 54.4%; Score 37; DB 1; Length 1033;
Best Local Similarity 50.0%; Pred. No. 74;
Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 QVRFSSDYTN 12
DB 894 QURVWEQDYN 905

RESULT 9
CBBA_BACUK
ID CBBA_BACUK STANDARD; PRT; 1169 AA.
AC Q45705;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Pesticidal crystal protein cry8Ba (Insecticidal delta-endotoxin
DE CryVIIIb(a)) (Crystalline entomocidal protoxin) (134 kDa crystal
DE protein).
GN CRY8BA OR CRYVIIIb(A) OR 50C(B).
OS Bacillus thuringiensis (subsp. kumamotoensis).
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=132267;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NRL B-18746 / PS50C;
RA Michaels T.E., Foncecerra L., Narva K.E.;
RT "Process for controlling scarab pests with Bacillus thuringiensis
RT isolates.";
RL Patent number WO9315206, 05-AUG-1993.
CC -!- FUNCTION: PROMOTES COLLOIDSMOTIC LYSIS BY BINDING TO THE MIDGUT
CC EPITHELIAL CELLS OF INSECTS. ACTIVE ON VARIOUS SCARABEID BEETLES.
CC -!- DEVELOPMENTAL STAGE: The crystal protein is produced during
CC sporulation and is accumulated both as an inclusion and as part of
CC the spore coat.
CC -!- MISCELLANEOUS: Toxic segment of the protein is located in the N-
CC terminus.
CC -!- SIMILARITY: Belongs to the delta endotoxin family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; L04735; AAA20892.1;
CC PIR; A47704; A47704.
CC HSP; P26221; 1TF4.
CC InterPro; IPR001956; CBD 3.
CC InterPro; IPR008965; Cellul. bind.
CC InterPro; IPR001701; Glyco_hydro_9.
CC InterPro; IPR008928; Glyco_trans_4hp.
CC Pfam; PF00942; CBM 3; 2.
CC Pfam; PF00759; Glyco_hydro_9; 1.
CC ProDom; PD001947; CBD 3; 1.
CC PROSITE; PS00592; GLYCOSYL_HYDROL_F9_1; 1.
CC PROSITE; PS00698; GLYCOSYL_HYDROL_F9_2; 1.

```

```

DR EMBL; U04365; AAA21118.1;
DR HSP; P07130; 1DLC.
DR InterPro; IPR001178; Endotoxin.
DR InterPro; IPR005638; endotoxin_C.
DR InterPro; IPR005639; endotoxin_N.
DR InterPro; IPR008979; Gal_bind_like.
DR Pfam; PF00555; endotoxin_1.
DR Pfam; PF03944; endotoxin_C; 1.
DR Pfam; PF03945; endotoxin_N; 1.
DR Toxin; Sporulation.
SQ SEQUENCE 1169 AA; 133543 MW; 22EEFCF5BD699909 CRC64;

Query Match 54.4%; Score 37; DB 1; Length 1169;
Best Local Similarity 38.5%; Pred. No. 84;
Matches 5; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

QY 1 QVRFSSDYTN 13
DB 573 RIRIVASNETSY 585

RESULT 10
GUN1_CLOTM
ID GUN1_CLOTM STANDARD; PRT; 879 AA.
AC Q02934;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Endoglucanase I precursor (EC 3.2.1.4) (EGI) (Endo-1,4-beta-glucanase)
DE (Cellulase I).
GN CELI.
OS Clostridium thermocellum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1515;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 56-69.
RC STRAIN=NCIB 10682;
RX MEDLINE=93171873; PubMed=8436949;
RA Hazlewood G.P., Davidson K., Laurie J.I., Huskisson N.S.,
RA Gilbert H.J.;
RT "Gene sequence and properties of CELI, a family E endoglucanase from
RT Clostridium thermocellum.";
RL J. Gen. Microbiol. 139:307-316 (1993).
CC -!- FUNCTION: THIS ENZYME CATALYZES THE ENDOHYDROLYSIS OF 1,4-BETA-
CC GLUCOSIDIC LINKAGES IN CELLULOSE, LICHENIN AND CEREAL BETA-D-
CC GLUCANS. PRINCIPALLY ACTIVE AGAINST BARLEY BETA-GLUCAN.
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose, lichenin and cereal beta-D-glucans.
CC -!- PATHWAY: Cellulose degradation.
CC -!- SIMILARITY: Belongs to cellulase family E (family 9 of glycosyl
CC hydrolases).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; L04735; AAA20892.1;
CC PIR; A47704; A47704.
CC HSP; P26221; 1TF4.
CC InterPro; IPR001956; CBD 3.
CC InterPro; IPR008965; Cellul. bind.
CC InterPro; IPR001701; Glyco_hydro_9.
CC InterPro; IPR008928; Glyco_trans_4hp.
CC Pfam; PF00942; CBM 3; 2.
CC Pfam; PF00759; Glyco_hydro_9; 1.
CC ProDom; PD001947; CBD 3; 1.
CC PROSITE; PS00592; GLYCOSYL_HYDROL_F9_1; 1.
CC PROSITE; PS00698; GLYCOSYL_HYDROL_F9_2; 1.

```

```

KW Cellulose degradation; Hydrolase; Glycosidase; Signal.
FT CHAIN 1 55
FT CHAIN 56 879
FT DOMAIN 58 518
FT DOMAIN 729 879
FT ACT_SITE 448 448
FT ACT_SITE 486 486
FT ACT_SITE 495 495
SQ SEQUENCE 879 AA; 97796 MW; 35A60069A514A927 CRC64;

Query Match 53.7%; Score 36.5; DB 1; Length 879;
Best Local Similarity 50.0%; Pred. No. 76;
Matches 7; Conservative 5; Mismatches 1; Indels 1; Gaps 1;

QY 1 QVRVRES-SDYTNV 13
Db 839 EVQIRFSKEDWSNY 852

RESULT 11
GUNB_CALSA STANDARD; PRT; 1039 AA.
ID GUNB_CALSA
AC P10474;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Endoglucanase/exoglucanase B precursor [Includes: Endoglucanase
DE (EC 3.2.1.4) (Endo-1,4-beta-glucanase) (Cellulase)
DE (Cellulohydrolase); Exoglucanase (EC 3.2.1.1.91) (Exocellobiohydrolase)
DE (1,4-beta-cellobiohydrolase)].
GN CELB.
OS Caldocellum saccharolyticum (Caldicellulosiruptor saccharolyticus).
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Syntrophomonadaceae;
OC Caldicellulosiruptor.
OX NCBI_TaxID=44001;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=8909398; PubMed=2789517;
RA Saul D.J., Williams L.C., Love D.R., Chamley I.W., Bergquist P.L.;
RT "Nucleotide sequence of a gene from Caldocellum saccharolyticum
RT encoding for exocellulase and endocellulase activity.";
RL Nucleic Acids Res. 17:439-439(1989).
CC -!- FUNCTION: THIS PROTEIN IS MADE UP OF TWO DOMAINS: THE N-TERMINAL
CC DOMAIN HAS EXOGLUCANASE ACTIVITY WHILE THE C-TERMINAL DOMAIN IS
CC AN ENDOGLUCANASE.
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose, lichenin and cereal beta-D-glucans.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-glucosidic linkages
CC in cellulose and cellotetraose, releasing cellobiose from the non-
CC reducing ends of the chains.
CC -!- SIMILARITY: In the N-terminal section; belongs to cellulase family
CC F (family 10 of glycosyl hydrolases).
CC -!- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO CELLULASE FAMILY
CC A (FAMILY 5 OF GLYCOSYL HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X13602; CAA31936.1; -.
CC PIR; S02711; S02711.
CC HSP; Q06851; INRC.
CC InterPro; IPR001956; CBD 3.
CC InterPro; IPR008965; Cellul. bind.
CC InterPro; IPR001000; Glyco_hydro_10.
CC InterPro; IPR001547; Glyco_hydro_5.
CC Pfam; PF00942; CM 3; 1.
CC Pfam; PF00150; cellulase; 1.
CC Pfam; PF00331; Glyco_hydro_10; 1.

PRINTS; PR00134; GLHYDLASE10.
PRODOM; PD001947; CBD_3; 1.
SMART; SM00633; Glyco_10; 1.
PROSITE; PS00591; GLYCOSYL_HYDROL_F10; 1.
PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
Cellulose degradation; Hydrolase; Glycosidase; Repeat;
Multifunctional enzyme; Signal.
FT SIGNAL 1 28
FT CHAIN 29 1039
FT DOMAIN 376 416
FT DOMAIN 417 570
FT DOMAIN 571 618
FT ACT_SITE 177 177
FT ACT_SITE 285 285
FT ACT_SITE 792 792
FT ACT_SITE 1039 AA; 117641 MW; 0E0378171594DDAE CRC64;
SQ SEQUENCE 1039 AA; 117641 MW; 0E0378171594DDAE CRC64;

Query Match 53.7%; Score 36.5; DB 1; Length 1039;
Best Local Similarity 42.9%; Pred. No. 91;
Matches 6; Conservative 7; Mismatches 0; Indels 1; Gaps 1;

QY 1 QVRVRES-SDYTNV 13
Db 522 EIQIRFNKSDWSNY 535

RESULT 12
MANB_CALSA STANDARD; PRT; 1331 AA.
ID MANB_CALSA
AC P22533;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Beta-mannanase/endoglucanase A precursor [Includes: Mannan endo-1,4-
DE beta-mannosidase A (EC 3.2.1.78) (Beta-mannanase) (Endo-1,4-
DE mannanase); Endo-1,4-beta-glucanase (EC 3.2.1.4) (Cellulase)].
GN MANA.
OS Caldocellum saccharolyticum (Caldicellulosiruptor saccharolyticus).
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Syntrophomonadaceae;
OC Caldicellulosiruptor.
OX NCBI_TaxID=44001;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9311939; PubMed=1476429;
RA Gibbs M.D., Saul D.J., Luthi E., Bergquist P.L.;
RT "The beta-mannanase from 'Caldocellum saccharolyticum' is part of a
RT multidomain enzyme.";
RL Appl. Environ. Microbiol. 58:3864-3867(1992).
RN [2]
RP SEQUENCE OF 1-346 FROM N.A.
RX MEDLINE=91247819; PubMed=2039230;
RA Luehti E., Jasmat N.B., Grayling R.A., Love D.R., Bergquist P.L.;
RT "Cloning, sequence analysis, and expression in Escherichia coli of a
RT gene coding for a beta-mannanase from the extremely thermophilic
RT bacterium 'Caldocellum saccharolyticum'.";
RL Appl. Environ. Microbiol. 57:694-700(1991).
CC -!- FUNCTION: DEGRADATION OF HEMICELLULOSES. THE SECOND MOST ABUNDANT
CC POLYSACCHARIDES IN NATURE. CONTAINS TWO CATALYTIC DOMAINS WITH
CC MANNANASE AND ENDOGLUCANASE ACTIVITIES.
CC -!- CATALYTIC ACTIVITY: Random hydrolysis of 1,4-beta-D-mannosidic
CC linkages in mannans, galactomannans, glucomannans, and
CC galactoglucanans.
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose, lichenin and cereal beta-D-glucans.
CC -!- MISCELLANEOUS: This enzyme is most active at pH 6 and 80 degrees
CC Celsius.
CC -!- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO CELLULASE FAMILY
CC A (FAMILY 5 OF GLYCOSYL HYDROLASES).
CC -!- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO CELLULASE FAMILY
CC J (FAMILY 44 OF GLYCOSYL HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

```

CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----  
 DR EMBL; L01257; AAA71887.1; -  
 DR EMBL; M36063; AAA72861.1; -  
 DR FIR; A48954; A48954.  
 DR HSSP; Q06851; INEC.  
 DR InterPro; IPR001956; CBD 3.  
 DR InterPro; IPR008965; Cellul\_bind.  
 DR InterPro; IPR001547; Glyco\_Hydro\_5.  
 DR Pfam; PF00942; CBM 3; 2.  
 DR Pfam; PF00150; Cellulase; 1.  
 DR ProDom; PD001947; CBD 3; 2.  
 DR PROSITE; PS00659; GLYCOSYL\_HYDROL\_F5; 1.  
 DR Hydrolase; Glycosidase; Cellulose degradation; Signal;  
 KW Multifunctional enzyme.  
 FT SIGNAL 1 41  
 FT CHAIN 42 1331  
 FT DOMAIN 42 325  
 FT DOMAIN 326 361  
 FT DOMAIN 362 518  
 FT DOMAIN 519 564  
 FT DOMAIN 565 720  
 FT DOMAIN 721 780  
 FT DOMAIN 781 1331  
 FT ACT\_SITE 182 182  
 FT ACT\_SITE 257 257  
 FT CONFLICT 338 338  
 FT CONFLICT 340 346  
 FT CONFLICT 1331 AA; 146892 MW; PFBCA51BB8DF0E0 CRC64;  
 SQ SEQUENCE 1331 AA; 146892 MW; PFBCA51BB8DF0E0 CRC64;  
 Query Match 53.7%; Score 36.5; DB 1; Length 1331;  
 Best Local Similarity 42.9%; Pred. No. 1.2e+02;  
 Matches 6; Conservative 7; Mismatches 0; Indels 1; Gaps 1;  
 QY 1 QVRVRS-SDYTNV 13  
 Db 670 EIQRFNKSDNSNY 683  
 RESULT 13  
 LG27\_EUBNA  
 ID LG27\_EUBNA STANDARD; PRT; 71 AA.  
 AC P40654;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE SRI-related protein LG27 (fragment).  
 OS Eubapharis macularius (Leopard gecko).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Lepidosauria; Squamata; Scleroglossa; Gekkota; Eublepharidae;  
 CC Eublepharus.  
 CC NCBI\_TaxID=8564;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93184703; PubMed=8443573;  
 RA Coriati A.M., Mueller U., Harry J.L., Uwanogho D., Sharpe P.T.;  
 RT "PCR amplification of SRY-related gene sequences evolutionary  
 conservation of the SRY-box motif";  
 RL PCR Methods Appl. 2:218-222(1993).  
 CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).  
 CC -!- SIMILARITY: Contains 1 HMG box domain.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; M86335; AAA92978.1; -  
 DR PIR; I50541; I50541.  
 DR HSSP; P48436; ISX9.  
 DR InterPro; IPR000910; HMG\_12\_box.  
 DR Pfam; PF00505; HMG\_box; 1.  
 DR SMART; SM00398; HMG; 1.  
 DR PROSITE; PS0118; HMG\_BOX 2; 1.  
 KW DNA-binding; Nuclear protein.  
 FT NON\_TER 1 1  
 FT DNA\_BIND 1 68  
 FT NON\_TER 71 71  
 FT HMG\_BOX.  
 SQ SEQUENCE 71 AA; 8919 MW; 23D2076B6C29F71A CRC64;  
 Query Match 52.9%; Score 36; DB 1; Length 71;  
 Best Local Similarity 50.0%; Pred. No. 6.6;  
 Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;  
 QY 2 VVRVPSDYTNV 13  
 Db 56 VRLXWADYFNY 67  
 RESULT 14  
 RS8\_ARCFU  
 ID RS8\_ARCFU STANDARD; PRT; 131 AA.  
 AC O28369;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE 30S ribosomal protein S8p.  
 GN RPS8P OR AFI910.  
 OS Archaeoglobus fulgidus.  
 CC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;  
 CC Archaeoglobaceae; Archaeoglobus.  
 CC NCBI\_TaxID=2234;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN=VC-16 / DSM 4304 / ATCC 49558;  
 RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,  
 RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,  
 RA Richardson D.L., Kerlavage A.R., Graham D.E., Kyrpides N.C.,  
 RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,  
 RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,  
 RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,  
 RA Overbeek R., Cockayne J.D., Weidman J.F., McDonald L., Utterback T.,  
 RA Cotton M.D., Spriggs T., Artach P., Kaine B.P., Sykes S.M.,  
 RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,  
 RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,  
 RA Venter J.C.;  
 RT "The complete genome sequence of the hyperthermophilic, sulphate-  
 reducing archaeon Archaeoglobus fulgidus";  
 RL Nature 390:364-370(1997).  
 CC -!- FUNCTION: One of the primary rRNA binding proteins, it binds  
 CC directly to 16S rRNA central domain where it helps coordinate  
 CC assembly of the platform of the 30S subunit (by similarity).  
 CC -!- SUBUNIT: Part of the 30S ribosomal subunit.  
 CC -!- SIMILARITY: Belongs to the S8p family of ribosomal proteins.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----  
 DR EMBL; AE000971; AAB89341.1; -  
 DR FIR; E69488; E69488.  
 DR HSSP; P56209; 1SE1.  
 DR TIGR; AF1910; -  
 DR HAMAP; MF\_01302; -; 1.

DR InterPro: IPR000630; Ribosomal\_S8.  
DR Pfam: PFC0410; Ribosomal\_S8; 1.  
DR ProDom: PD001098; Ribosomal\_S8; 1.  
DR PROSITE: PS00053; RIBOSOMAL\_S8; 1.  
DR KEGG: RIBOSOMAL protein; RNA-binding; rRNA-binding; Complete proteome.  
DR SWISS-PROT: S31; Ribosomal protein S31; mitochondrial precursor (S31mt) (MRP-S31)  
SEQUENCE 131 AA; 14684 MW; E1B413B376449B3E CRC64;

Query Match 52.9%; Score 36; DB 1; Length 131;  
Best Local Similarity 58.3%; Pred.No. 13;  
Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 VVRVFSFSDVTNY 13  
          :| |||| |  
DB 76 IRPFSSSVTEY 87

RESULT 15  
ID RT31 HUMAN STANDARD; PRT; 395 AA.  
IT ID RT31 HUMAN STANDARD; PRT; 395 AA.  
AC Q92665; O8WTV8;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE 28S ribosomal protein S31, mitochondrial precursor (S31mt) (MRP-S31)  
DE (Imogen 38).  
DE MRPS31 OR IMOGN38.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
NCBI\_Taxid=9606;  
RN [1]  
RN SEQUENCE FROM N.A.  
RC TISSUE=Skin;  
RC Hutton J.C., Roep B.O.;  
RA "Human Imogen 38". T-cell and antibody responses in newly diagnosed  
RT diabetic subjects.;  
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RN SEQUENCE FROM N.A.  
RC TISSUE=Testis;  
RC MEDLINE=22388257; PubMed=12477932;  
RX Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buettow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Ustin T.B., Toshiruki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
RA Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S.C., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalón D.K., Wuzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Kettelman M., Madan A.C., Rodrigues S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalley D.E.,  
RA Schnurch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length  
human and mouse cDNA sequences";  
Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
RN [3]  
RN IDENTIFICATION.  
RX MEDLINE=21276436; PubMed=11279123;  
RX Koc E.C., Burkhardt W., Blackburn K., Moseley A., Sprengel L.L.;  
RT "The small subunit of the mammalian mitochondrial ribosome:  
RT identification of the full complement of ribosomal proteins present.";  
RL J. Biol. Chem. 276:19363-19374(2001).  
CC -I- SUBUNIT: Component of the mitochondrial ribosome small subunit  
CC (28S) which comprises a 12S rRNA and about 30 distinct proteins.  
CC -I- SUBCELLULAR LOCATION: Mitochondrial.  
CC This SWISS-PROT entry is copyright It is produced through a collaboration

This Page Blank (uspto)

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2004, 07:34:11 ; Search time 6.8817 Seconds  
(without alignments)  
596.081 Million cell updates/sec

Title: US-09-805-290A-11  
Perfect score: 68  
Sequence: 1 QVRVRFSSDYTNV 13

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPREMBL 25:\*

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phage:\*
- 10: sp\_plant:\*
- 11: sp\_rodent:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_virus:\*
- 16: sp\_bacteriopl:\*
- 17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	43	63.2	513	16	Q85E58
2	43	63.2	513	16	Q8DZQ0
3	43	63.2	1129	16	Q87PU3
4	41	60.3	773	16	Q88DY5
5	40	58.8	152	16	Q8YMP7
6	40	58.8	346	16	Q8FMT6
7	40	58.8	615	10	Q81985
8	40	58.8	1067	16	Q8A3P9
9	40	58.8	3130	12	Q9VZN8
10	40	58.8	3132	12	Q9QN08
11	39	57.4	96	13	Q9DG43
12	39	57.4	336	16	Q8F9G1
13	39	57.4	352	4	Q8T834
14	39	57.4	407	16	Q8V875
15	39	57.4	469	5	Q9VF93
16	39	57.4	470	5	Q8T072

#### ALIGNMENTS

##### RESULT 1

Q85E58 PRELIMINARY; PRT; 513 AA.

AC Q85E58; 01-MAR-2003 (TRENBLrel. 23, Created)

DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)

DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)

DE Hypothetical protein.

GN GBS1084.

OS Streptococcus agalactiae (serotype III).

OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;

OC Streptococcus.

OX NCBI\_TaxID=216495;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=NEM316 / Serotype III;

RX MEDLINE=22242508; PubMed=12354221;

RA Glaser P., Rusniok C., Buchrieser C., Chevalier F., Frangeul L., Msadek T., Zouine M., Couve E., Lalloui L., Poyart C., Trieu-Cuot P., Kunst F.;

RT "Genome sequence of Streptococcus agalactiae, a pathogen causing invasive neonatal disease.";

RL Mol. Microbiol. 45:1499-1513(2002).

DR EMBL; AL766845; CAD46743.1; --

DR Sagaidist; gbs1084; --

DR GO; GO:0016020; C:membrane; IEA.

DR GO; GO:0005524; F:ATP binding; IEA.

DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. .; IEA.

DR GO; GO:0000166; F:nucleotide binding; IEA.

DR GO; GO:0006810; P:transport; IEA.

DR InterPro; IPR003593; AAA\_ATPase.

DR InterPro; IPR003439; ABC\_transporter.

DR Pfam; PF00005; ABC\_tran; 2.

DR ProDom; PD000006; ABC\_transporter; 1.

DR SMART; SM00382; AAA; 2.

DR PROSITE; PS00211; ABC\_TRANSPORTER\_1; 2.

DR PROSITE; PS00893; ABC\_TRANSPORTER\_2; 2.

KW Hypothetical protein; Complete proteome.

SQ SEQUENCE 513 AA; 58190 MW; BAZCS96BC5983C88 CRC64;

Q8BZQ8 mus musculus

Q81R04 drosophila

Q8VX90 drosophila

Q75182 homo sapien

Q8RPL7 anaplasma p

Q8P935 methanosarc

Q8X683 escherichia

Q88XJ6 lactobacilli

Q8YV54 arabidopsis

Q8VVS9 drosophila

Q8NZ14 drosophila

Q8D7L4 vibrio vuln

Q8GA49 escherichia

Q81C23 plasmodium

Q62178 caenorhabdi

Q93529 caenorhabdi

Q9GM64 bos taurus

Q30232 bison bison

Q8GUL3 bos taurus

Q97M13 clostridium

Q8EU01 oceanobacil

Q8P336 methanosarc

Q58769 pyrococcus

Q7TX4 mus musculus

Q67645 gallid herp

Q8XX29 ralsconia s

Q8X9U1 escherichia

Q8FHM7 escherichia

Q83R82 shigella fl

```

Query Match          63.2%; Score 43; DB 16; Length 513;
Best Local Similarity 70.0%; Pred. No. 24;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 4 VRFSSDYNNY 13
DB 237 VRYSGDYNNF 246

RESULT 2
Q8DZQO
ID Q8DZQO PRELIMINARY; PRT; 513 AA.
AC Q8DZQO;
DT 01-WAR-2003 (TREMELrel. 23, Created)
DT 01-VAR-2003 (TREMELrel. 23, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE ABC transporter, ATP-binding protein.
GN SAG1049.
OS Streptococcus agalactiae (serotype V).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=216466;
RN [1]_
RP SEQUENCE FROM N.A.
RC STRAIN=2603 V/R / Serotype V;
RX MEDLINE=2222988; PubMed=12200547;
RA Tettelin H., Masignani V., Cieslewicz M.J., Eisen J.A., Peterson S.,
RA Wessels M.R., Paulsen I.T., Nelson K.E., Margarit I., Read T.D.,
RA Madoff L.C., Wolf A.M., Beanan M.J., Brinkac L.M., Daugherty S.C.,
RA DeBoy R.T., Durkin A.S., Kolonay J.F., Madupu R., Lewis M.R.,
RA Radune D., Fedorova N.B., Scanlan D., Khouiri H., Mulligan S.,
RA Carthy H.A., Cline R.T., Van Aken S.E., Gill J., Scarselli M., Mora M.,
RA Iacobini E.T., Brettoni C., Galli G., Mariani M., Vegni F., Malone D.,
RA Rinaldo D., Rappuoli R., Telford J.L., Kasper D.L., Grandi G.,
RA Fraser C.M.;
RT "Complete genome sequence and comparative genomic analysis of an
RT emerging human pathogen, serotype V Streptococcus agalactiae."
RL Proc. Natl. Acad. Sci. U.S.A. 99:12391-12396 (2002).
DR EMBL; AF014240; AAM99931.1; -.
DR TIGR; SAG1049; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti...; IEA.
DR GO; GO:0000166; F:nucleotide binding; IEA.
DR GO; GO:0008810; F:transport; IEA.
DR InterPro; IPR003593; AAA ATPase.
DR Pfam; PF00005; ABC tran; 2.
DR ProDom; PD000006; ABC transporter; 1.
DR SMART; SM00382; AAA; 2.
DR PROSITE; PS00211; ABC_TRANSPORTER_1; 2.
DR PROSITE; PS00893; ABC_TRANSPORTER_2; 2.
KW ATP-binding; Complete proteome.
SQ SEQUENCE 513 AA; 58191 MW; 6F9C7BE5C4A10392 CRC64;

Query Match          63.2%; Score 43; DB 16; Length 513;
Best Local Similarity 70.0%; Pred. No. 24;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 4 VRFSSDYNNY 13
DB 237 VRYSGDYNNF 246

RESULT 3
Q87PU3
ID Q87PU3 PRELIMINARY; PRT; 1129 AA.
AC Q87PU3;
DT 01-JUN-2003 (TREMELrel. 24, Created)
DT 01-JUN-2003 (TREMELrel. 24, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE Putative tcmF-related protein.

```

```

GN VP1408.
OS Vibrio parahaemolyticus.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=670;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RIMD 2210633 / Serotype O3:K6;
RX MEDLINE=22508454; PubMed=12620739;
RA Makino K., Oshima K., Kurokawa K., Yokoyama K., Uda T., Tagomori K.,
RA Iijima Y., Najima M., Nakano M., Yamashita A., Kubota Y., Kimura S.,
RA Yasunaga T., Honda T., Shinagawa H., Hattori M., Iida T.;
RT "Genome sequence of Vibrio parahaemolyticus: a pathogenic mechanism
RT distinct from that of V. cholerae."
RL Lancet 361:743-749 (2003).
DR EMBL; AP005078; BAC59671.1; -.
DR GO; GO:0006364; P:RNA processing; IEA.
DR InterPro; IPR000238; Rib_bind_factA.
DR PROSITE; PS01319; RBFA; 1.
KW Complete proteome.
SQ SEQUENCE 1129 AA; 128541 MW; 7D12799A09BA89D9 CRC64;

Query Match          63.2%; Score 43; DB 16; Length 1129;
Best Local Similarity 53.8%; Pred. No. 55;
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 QVRFFSSDYNNY 13
DB 698 QVRKKYTDYNNY 710

RESULT 4
Q88DY5
ID Q88DY5 PRELIMINARY; PRT; 773 AA.
AC Q88DY5;
DT 01-JUN-2003 (TREMELrel. 24, Created)
DT 01-JUN-2003 (TREMELrel. 24, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE Penicillin-binding protein.
GN MRCB OR PP4683.
OS Pseudomonas putida (strain KT2440).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=160488;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22423060; PubMed=12534463;
RA Nelson K.E., Weinel C., Paulsen I.T., Dodson R.J., Hilbert H.,
RA Martins dos Santos V.A.P., Fouts D.E., Gill S.R., Pop M., Holmes M.,
RA Brinkac L., Beanan M., DeBoy R.T., Daugherty S., Kolonay J.,
RA Madupu R., Nelson W., White O., Peterson J., Khouiri H., Hance I.,
RA Chris Lee P., Holtzapple E., Scanlan D., Tran K., Moazzes A.,
RA Utterback T., Rizzo M., Lee K., Kosack D., Moescl D., Wedler H.,
RA Lauber J., Stjepandic D., Hoheisel J., Straetz M., Heim S.,
RA Kiewitz C., Eisen J., Tirmis K.N., Duesterhoeft A., Tuemmler B.,
RA Fraser C.M.;
RT "Complete genome sequence and comparative analysis of the
RT metabolically versatile Pseudomonas putida KT2440."
RL Environ. Microbiol. 4:799-808 (2002).
DR EMBL; AF016791; AAN70256.1; -.
DR TIGR; PP4683; -.
DR GO; GO:0003274; C:cell wall (sensu Bacteria); IEA.
DR GO; GO:0003824; F:catalytic activity; IEA.
DR GO; GO:0008658; F:penicillin binding; IEA.
DR GO; GO:0009252; P:peptidoglycan biosynthesis; IEA.
DR InterPro; IPR001264; Glyco_trans_51.
DR InterPro; IPR001460; Transpeptidase.
DR Pfam; PF00912; Transglycosyl; 1.
DR Pfam; PF00905; Transpeptidase; 1.
KW Complete proteome.
SQ SEQUENCE 773 AA; 85156 MW; 3ED94FF21E0C6AFE CRC64;

Query Match          60.3%; Score 41; DB 16; Length 773;

```



Best Local Similarity 98.9%; Pred. No. 85;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 VVRVFSDDY 10  
DB 127 VVRVFSDDY 135

RESULT 5  
Q8YMP7 PRELIMINARY; PRT; 152 AA.  
AC Q8YMP7;  
DT 01-MAR-2002 (TREMELrel. 20, Created)  
DT 01-MAR-2002 (TREMELrel. 20, Last sequence update)  
DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)  
DE Hypothetical protein A14886.  
GN A14886.  
OS Anabaena sp. (strain PCC 7120).  
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.  
OX NCBI\_TaxID=103690;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21595285; PubMed=11759840;  
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,  
RA Watanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,  
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,  
RA Nakazaki N., Shimpo S., Sugimoto M., Takazawa M., Yamada M.,  
RA Yasuda M., Tabata S.;  
RT "Complete genomic sequence of the filamentous nitrogen-fixing  
RT cyanobacterium Anabaena sp. strain PCC 7120.";  
RL DNA Res. 8:205-213(2001).  
DR EMBL; AP003597; BAB76585.1; -;  
DR PIR; AF2416; AF2416.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 152 AA; 17008 MW; 5376E02EFCFDB4E5 CRC64;

Query Match 58.8%; Score 40; DB 16; Length 152;  
Best Local Similarity 54.5%; Pred. No. 23;  
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 3 VVRVFSDDY 13  
DB 82 RIRFTGDIYSY 92

RESULT 6  
Q8FMT6 PRELIMINARY; PRT; 346 AA.  
AC Q8FMT6;  
DT 01-MAR-2003 (TREMELrel. 23, Created)  
DT 01-MAR-2003 (TREMELrel. 23, Last sequence update)  
DT 01-MAR-2003 (TREMELrel. 23, Last annotation update)  
DE Hypothetical protein.  
GN C22413.  
OS Corynebacterium efficiens.  
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
OC Corynebacterineae; Corynebacteriaceae; Corynebacterium.  
OX NCBI\_TaxID=152794;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=YS-314 / AJ 12310 / DSM 44549 / JCM 11189;  
RA Kawarabayashi Y., Yamazaki J., Hino Y., Kikuchi H., Nakamura Y.,  
RA Ikeo K., Suzuki M., Maehina J., Itoh T., Yamagishi A., Nishio Y.,  
RA Ueda Y., Sugimoto S.;  
RT "The entire genomic sequence of Corynebacterium efficiens YS-314.";  
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AP005222; BAC19223.1; -;  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 346 AA; 36872 MW; 9243CB84E9EF6049 CRC64;

Query Match 58.8%; Score 40; DB 16; Length 346;  
Best Local Similarity 63.6%; Pred. No. 55;  
Matches 7; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 VVRVFSDDY 13  
DB 132 REREGDY 142

RESULT 7  
O81985 PRELIMINARY; PRT; 615 AA.  
AC O81985;  
DT 01-NOV-1998 (TREMELrel. 08, Created)  
DT 01-NOV-1998 (TREMELrel. 08, Last sequence update)  
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)  
DE 1,2-beta-fructan 1F-fructosyltransferase precursor  
DE EC 2.4.1.100.  
GN FFT-1.  
OS Helianthus tuberosus (Jerusalem artichoke).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;  
OC campanulids; Asterales; Asteraceae; Asteroidae; Heliantheae;  
OC Helianthus.  
OX NCBI\_TaxID=4233;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=cv. Colombia; TISSUE=Tuber;  
RX MEDLINE=98426382; PubMed=9753774;  
RA van der Meer I.M., Koops A.J., Hakkert J.C., van Tunen A.J.;  
RT "Cloning of the fructan biosynthesis pathway of Jerusalem artichoke.";  
RL Plant J. 15:489-500(1998).  
CC -1- SIMILARITY: BELONGS TO FAMILY 32 OF GLYCOSYL HYDROLASES.  
DR EMBL; AJ009756; CAA08811.1; -;  
DR GO; GO:0047207; F:1.2-beta-fructan 1F-fructosyltransferase ac. . . ; IEA.  
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . . ; IEA.  
DR GO; GO:0016757; F:transferase activity, transferring glycosyl. . . ; IEA.  
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.  
DR InterPro; IPR001362; Glyco\_hydro\_32.  
DR Pfam; PF00251; Glyco\_hydro\_32; 1.  
DR SMART; SM00640; Glyco\_32; 1.  
KW Glycoprotein; Glycosidase; Glycosyltransferase; Hydrolase; Signal;  
KW Transferase.  
FT SIGNAL 1 79 POTENTIAL.  
FT CHAIN 80 615 1,2-BETA-FRUCTAN 1F-FRUCTOSYLTRANSFERASE.  
SQ SEQUENCE 615 AA; 69258 MW; F4D746595546DB3 CRC64;

Query Match 58.8%; Score 40; DB 10; Length 615;  
Best Local Similarity 58.3%; Pred. No. 1e+02;  
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 2 VVRVFSDDY 13  
DB 252 VLVYTTDTNY 263

RESULT 8  
Q8A3P9 PRELIMINARY; PRT; 1067 AA.  
ID Q8A3P9  
AC Q8A3P9;  
DT 01-JUN-2003 (TREMELrel. 24, Created)  
DT 01-JUN-2003 (TREMELrel. 24, Last sequence update)  
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)  
DE Putative outer membrane protein, probably involved in nutrient  
DE binding.  
GN B12905.  
OS Bacteroides thetaiotaomicron.  
OC Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;  
OC Bacteroidaceae; Bacteroides.  
OX NCBI\_TaxID=818;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=VPI-5482 / ATCC 29148;  
RX MEDLINE=22550858; PubMed=12663928;  
RA Xu J., Bjursell M.K., Himrod J., Deng S., Carmichael L.K.,  
RA Chiang H.C., Hooper L.V., Gordon J.I.;

RT "A genomic view of the human-Bacteroides thetaiotaomicron symbiosis."  
 RL Science 299:2074-2076(2003).

DR EMBL; AB016938; AAO78011.1; -;  
 DR GO; GO:0005509; F:calcium ion binding; IEA.  
 DR GO; GO:0004784; P:superoxide dismutase activity; IEA.  
 DR GO; GO:0006801; P:superoxide metabolism; IEA.  
 DR InterPro; IPR008969; Carboxypeptid\_reg.  
 DR InterPro; IPR002048; EF-hand.  
 DR InterPro; IPR001189; SODismutase.  
 DR ProDom; PD000475; SODismutase; 1.  
 DR PROSITE; PS00018; EF\_HAND; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 1067 AA; 119530 MW; 9B3DEAF654C968A CRC64;

Query Match 58.8%; Score 40; DB 16; Length 1067;  
 Best Local Similarity 46.2%; Pred. No. 1.8e+02;  
 Matches 6; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 1 QVVRFSDDYNY 13  
 DB 400 EVKVRQADFDY 412

RESULT 9

ID Q9YZN8 PRELIMINARY; PRT; 3130 AA.  
 AC Q9YZN8;  
 DT 01-MAY-1999 (TREMBLrel. 10, Created)  
 DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)  
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
 DE Polyprotein.  
 OS Japanese Yam mosaic virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Potyviridae;  
 OC Potyvirus.  
 OC NCBI\_TaxID=79917;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Japanese Yam 1;  
 RX MEDLINE=99399236; PubMed=10470250;  
 RA Fuji S., Nakamae H.;  
 RT "Complete nucleotide sequence of the genomic RNA of a Japanese Yam mosaic virus, a novel potyvirus in Japan."  
 RL Arch. Virol. 144:231-240(1999).  
 CC -!- SIMILARITY: BELONGS TO THE POTYVIRUSES POLYPROTEIN FAMILY.  
 DR EMBL; AB016500; BAA36278.1; -;  
 DR MEROPS; C04.001; -;  
 DR MEROPS; C06.001; -;  
 DR MEROPS; S30.001; -;  
 DR GO; GO:0019028; C:Viral capsid; IEA.  
 DR GO; GO:0005524; F:ATP binding; IEA.  
 DR GO; GO:0008026; F:ATP dependent helicase activity; IEA.  
 DR GO; GO:0004197; F:cysteine-type endopeptidase activity; IEA.  
 DR GO; GO:0016787; F:hydrolase activity; IEA.  
 DR GO; GO:0003676; F:nucleic acid binding; IEA.  
 DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.  
 DR GO; GO:0005198; F:structural molecule activity; IEA.  
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR GO; GO:0006350; P:transcription; IEA.  
 DR GO; GO:0019079; P:viral genome replication; IEA.  
 DR InterPro; IPR001410; DEAD.  
 DR InterPro; IPR001650; Helicase\_C.  
 DR InterPro; IPR001730; Peptidase\_C4.  
 DR InterPro; IPR001456; Peptidase\_C6.  
 DR InterPro; IPR001592; Poty\_coat.  
 DR InterPro; IPR002540; Poty\_P1.  
 DR InterPro; IPR007095; RNA pol\_DS\_PS.  
 DR InterPro; IPR001205; RNA pol\_P3D.  
 DR InterPro; IPR007094; RNA pol\_PSVir.  
 DR InterPro; IPR006186; T\_phase\_apah.  
 DR Pfam; PF00271; helicase\_C; 1.  
 DR Pfam; PF00863; Peptidase\_C4; 1.  
 DR Pfam; PF00851; Peptidase\_C6; 1.  
 DR Pfam; PF00767; Poty\_coat; 1.

DR Pfam; PF01577; Poty\_P1; 1.  
 DR Pfam; PF00680; RNA\_dep\_RNA\_pol; 1.  
 DR PRINTS; PR00966; NIAPOTPTASE.  
 DR SMART; SM00487; DEXdc; 1.  
 DR SMART; SM00490; HELICc; 1.  
 DR PROSITE; PS00125; SER\_THR\_PHOSPHATASE; 1.  
 KW ATP-binding; Coat protein; Helicase; Hydrolase.  
 SQ SEQUENCE 3130 AA; 356794 MW; 0284FE17A7C5D0A CRC64;

Query Match 58.8%; Score 40; DB 12; Length 3130;  
 Best Local Similarity 70.0%; Pred. No. 5.7e+02;  
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 VVRFSDDYNY 11  
 DB 1775 IRARFSSDHT 1784

RESULT 10

ID Q9QN08 PRELIMINARY; PRT; 3132 AA.  
 AC Q9QN08;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
 DE Polyprotein.  
 OS Japanese Yam mosaic virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Potyviridae;  
 OC Potyvirus.  
 OC NCBI\_TaxID=79917;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=mild;  
 RA Fuji S., Nakamae H.;  
 RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=mild;  
 RX MEDLINE=20254545; PubMed=10795528;  
 RA Fuji S., Nakamae H.;  
 RT "Complete nucleotide sequence of the genomic RNA of a mild strain of Japanese Yam mosaic potyvirus."  
 RL Arch. Virol. 145:635-640(2000).  
 CC -!- SIMILARITY: BELONGS TO THE POTYVIRUSES POLYPROTEIN FAMILY.  
 DR EMBL; AB027007; BAA86288.1; -;  
 DR MEROPS; C04.001; -;  
 DR MEROPS; C06.001; -;  
 DR MEROPS; S30.001; -;  
 DR GO; GO:0019028; C:Viral capsid; IEA.  
 DR GO; GO:0005524; F:ATP binding; IEA.  
 DR GO; GO:0008026; F:ATP dependent helicase activity; IEA.  
 DR GO; GO:0004197; F:cysteine-type endopeptidase activity; IEA.  
 DR GO; GO:0016787; F:hydrolase activity; IEA.  
 DR GO; GO:0003676; F:nucleic acid binding; IEA.  
 DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.  
 DR GO; GO:0005198; F:structural molecule activity; IEA.  
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR GO; GO:0006350; P:transcription; IEA.  
 DR GO; GO:0019079; P:viral genome replication; IEA.  
 DR InterPro; IPR001410; DEAD.  
 DR InterPro; IPR001650; Helicase\_C.  
 DR InterPro; IPR001730; Peptidase\_C4.  
 DR InterPro; IPR001456; Peptidase\_C6.  
 DR InterPro; IPR001592; Poty\_coat.  
 DR InterPro; IPR002540; Poty\_P1.  
 DR InterPro; IPR007095; RNA pol\_DS\_PS.  
 DR InterPro; IPR001205; RNA pol\_P3D.  
 DR InterPro; IPR007094; RNA pol\_PSVir.  
 DR InterPro; IPR006186; T\_phase\_apah.  
 DR Pfam; PF00271; helicase\_C; 1.  
 DR Pfam; PF00863; Peptidase\_C4; 1.  
 DR Pfam; PF00851; Peptidase\_C6; 1.  
 DR Pfam; PF00767; Poty\_coat; 1.

```
DR Pfam; PF01577; Poty PI; 1.
DR Pfam; PF00680; RNA dep RNA pol; 1.
DR PRINTS; PR00966; NIAPOTYPTASE.
DR SMART; SM00487; DEXDC; 1.
DR SMART; SM00490; HELIC; 1.
DR PROSITE; PS00125; SER_THR_PHOSPHATASE; 1.
KW ATP-binding; Coat protein; Helicase; Hydrolase.
SQ SEQUENCE 3132 AA; 356199 MW; 2597B7197CFD47C7 CRC64;

Query Match 58.8%; Score 40; DB 12; Length 3132;
Best Local Similarity 70.0%; Pred. No. 5.7e+02;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 VVRVRFSSDYT 11
Db 1777 IRARFSSDHT 1786

RESULT 11
Q9DG43 Q9DG43 PRELIMINARY; PRT; 96 AA.
AC Q9DG43;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Proglucagon (Fragment).
OS Ambloplites rupestris (Rock bass).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes; Percoidae;
OC Centrarchidae; Ambloplites.
OX NCBI_TaxID=109273;
RN [1]
RP SEQUENCE FROM N.A.
RA Al-Mahrouki A.A., Irwin D.M., Youson J.H.;
RT "Rock Bass Proglucagon.";
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF190499; AAC16778.1; -.
DR HSP; P01274; ICCN.
DR GO; GO:0005576; C:extracellular; IEA.
DR GO; GO:0005179; F:hormone activity; IEA.
DR InterPro; IPR00532; Glucagon.
DR Pfam; PF00123; hormone2; 2.
DR PRINTS; PR00275; GLUCAGON.
DR SMART; SM00070; GLUCA; 2.
DR PROSITE; PS00260; GLUCAGON; 1.
FT CHAIN 1
FT NON_TER 1
FT CHAIN 1 >29 GLUCAGON.
FT CHAIN 39 >70 GLUCAGON-LIKE PEPTIDE 1.
FT CHAIN 86 >96 GLUCAGON-LIKE PEPTIDE 2.
FT NON_TER 96
SQ SEQUENCE 96 AA; 11225 MW; 6435033EBDDC00CE CRC64;

Query Match 57.4%; Score 39; DB 13; Length 96;
Best Local Similarity 75.0%; Pred. No. 22;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 6 FSSDYTNY 13
Db 6 FTNDYNY 13

RESULT 12
Q8F9G1 Q8F9G1 PRELIMINARY; PRT; 336 AA.
AC Q8F9G1;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DE Hypothetical protein.
GN LA0234.
OS Leptospira interrogans.
OC Bacteria; Spirochaetes; Spirochaetales; Leptospiraceae; Leptospira.
```

```
OX NCBI_TaxID=173;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=56601 / Serogroup Icterohaemorrhagiae / Serovar lai;
RA Ren S.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE011212; AAN47433.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 336 AA; 39487 MW; F6A7E30E3FA978C9 CRC64;

Query Match 57.4%; Score 39; DB 16; Length 336;
Best Local Similarity 46.2%; Pred. No. 81;
Matches 6; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 1 QVRVRFSSDYTNY 13
Db 279 EIRFRFSGNSNY 291

RESULT 13
Q8TB34 Q8TB34 PRELIMINARY; PRT; 352 AA.
AC Q8TB34;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC Tissue=skin;
RA Strausberg R.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC025026; AAH25026.1; -.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR003822; PAH.
DR Pfam; PF02671; PAH; 3.
DR KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 352 AA; 39375 MW; 8BF091F27A4B8367 CRC64;

Query Match 57.4%; Score 39; DB 4; Length 352;
Best Local Similarity 53.8%; Pred. No. 85;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 QVRVRFSSDYTNY 13
Db 49 QVKIRFGSDPATY 61

RESULT 14
Q8Y875 Q8Y875 PRELIMINARY; PRT; 407 AA.
AC Q8Y875;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein lmo1042.
GN LMO1042.
OS Listeria monocytogenes.
OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
OX NCBI_TaxID=1639;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EGD-e / Serovar 1/2a;
RX MEDLINE=21537279; PubMed=11679669;
RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
RA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,
RA Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
```

RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,  
RA Entian K.-D., Fshih H., Garcia-del Portillo F., Garrido P.,  
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,  
RA Jones L.-M., Kaerst U., Krefit J., Kuhn M., Kunat F., Kurapkat G.,  
RA Madueno E., Maitouran A., Mata Vicente J., Ng E., Nedjari H.,  
RA Nordsiek G., Novella S., de Pablo B., Perez-Diaz J.-C., Purcell R.,  
RA Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,  
RA Vazquez-Boland J.-A., Voss H., Wenland J., Cossart P.,  
RA "Comparative genomics of *Listeria species*.";  
RL Science 294:849-852(2001).  
DR EMBL; AL591977; CAC99120.1; -.  
DR PIR; AB1205; AB1205.  
DR L1stList; LM001042; -.  
DR GO; GO:0004519; F:endonuclease activity; IEA.  
DR GO; GO:0003676; F:nucleic acid binding; IEA.  
DR GO; GO:0006777; F:Mo-molybdopter in cofactor biosynthesis; IEA.  
DR InterPro; IPR001504; Endonuclease.  
DR InterPro; IPR001453; MOCF biosynth.  
DR InterPro; IPR005111; MoeA C.  
DR InterPro; IPR005110; MoeA N.  
DR Pfam; PF00994; MOCF biosynth; 1.  
DR Pfam; PF03454; MoeA C; 1.  
DR Pfam; PF03453; MoeA N; 1.  
DR TIGRFAMs; TIGR00177; molyb syn; 1.  
DR PROSITE; PS01070; NUCLEASE NON SPEC; 1.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 407 AA; 45009 MW; C00F642F6966ADE6 CRC64;

Query Match 57.4%; Score 39; DB 16; Length 407;  
Best Local Similarity 63.6%; Pred. No. 99;  
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 QVVRFFSSDYT 11  
||| : : |||  
DB 330 QVRAKVASDYT 340

RESULT 15  
Q9VF93 ID Q9VF93 PRELIMINARY; PRT; 469 AA.  
AC Q9VF93;  
DT 01-MAY-2000 (TREMELrel. 13, Created)  
DT 01-MAY-2000 (TREMELrel. 13, Last sequence update)  
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)  
DE CG6276 protein.  
GN CG6276.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Berkley;  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams W.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berhan B.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,  
RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrieria S., Fleischmann W.,  
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,

RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Verkulov G., Milshina N.V., Mobarary C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissenbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of *Drosophila melanogaster*.";  
RL Science 287:2185-2195(2000).  
DR EMBL; AB003708; AAP55167.1; -.  
DR FlyBase; FBgn0038316; CG6276.  
DR GO; GO:0003677; F:DNA binding; IEA.  
DR InterPro; IPR004210; BESS motif.  
DR InterPro; IPR006578; MADF.  
DR Pfam; PF02944; BESS; 1.  
DR SMART; SM00595; MADF; 1.  
SQ SEQUENCE 469 AA; 52284 MW; 03BCD5B0C3A4CEEF CRC64;

Query Match 57.4%; Score 39; DB 5; Length 469;  
Best Local Similarity 50.0%; Pred. No. 1.2e+02;  
Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 2 VVRFFSSDYTN 13  
:|: :|||  
DB 81 LRISFRREYTN 92

Search completed: October 5, 2004, 08:13:18  
Job time : 10.8812 secs

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2004, 07:04:10 ; Search time 9.89043 Seconds  
(without alignments)  
371.381 Million cell updates/sec

Title: US-09-805-290A-11

Perfect score: 68

Sequence: 1 QVRVRFSSDYTYN 13

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_29Jan04:\*

- 1: Geneseq1980s:\*
- 2: Geneseq1990s:\*
- 3: Geneseq2000s:\*
- 4: Geneseq2001s:\*
- 5: Geneseq2002s:\*
- 6: Geneseq2003as:\*
- 7: Geneseq2003bs:\*
- 8: Geneseq2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	68	100.0	13	AAE10544	Aae10544 Llama spe
2	68	100.0	130	AAE10557	Aae10557 HPL inhib
3	43	63.2	293	ABP26437	Abp26437 Streptoco
4	43	63.2	513	ABP30089	Abp30089 Streptoco
5	42	61.8	208	AAO17579	Aao17579 M catarrh
6	41	60.3	117	ABP33318	Abp33318 Human ORF
7	41	60.3	476	AAV81486	Aav81486 Pseudomon
8	41	60.3	748	ABU40154	Abu40154 Protein e
9	40	58.6	615	AAW00456	Aaw00456 l-Fructan
10	40	58.8	3132	AABS1367	Aabs1367 Japanese
11	39	57.4	176	AAAG73797	Aag73797 Human col
12	39	57.4	407	ABBA47642	Abba47642 Listeria
13	39	57.4	407	ABU32421	Abu32421 Protein e
14	39	57.4	469	ABBA62215	Abba62215 Drosophil
15	39	57.4	849	ABG16347	Abg16347 Novel hum
16	39	57.4	1048	ABBA4201	Abba4201 Drosophil
17	39	57.4	1130	ABG12663	Abg12663 Novel hum
18	39	57.4	1131	ABG07281	Abg07281 Novel hum
19	38.5	56.6	52	AAAM90186	Aam90186 Human imm
20	38	55.9	176	AAU97601	Aau97601 Haemophil
21	38	55.9	511	ABBA60770	Abba60770 Drosophil
22	37	54.4	113	AAE10550	Aae10550 Llama spe
23	37	54.4	100	AAAB75586	Aab75586 Human sec
24	37	54.4	165	AAAG75631	Aag75631 Human col
25	37	54.4	362	AAW00635	Aaw00635 ILTV glyc

26	37	54.4	362	2	AAW06787	Aaw06787 ILTV glyc
27	37	54.4	362	5	ABG31711	Abg31711 Infectiou
28	37	54.4	479	4	AAU29333	Aau29333 Novel nar
29	37	54.4	479	4	AAAB74925	Aab74925 E. coli a
30	37	54.4	1012	4	ABG18198	Abg18198 Novel hum
31	37	54.4	1012	4	ABG28888	Abg28888 Novel hum
32	37	54.4	1169	2	AAW06417	Aaw06417 Antiscara
33	36	52.9	41	5	ABP30755	Abp30755 Streptoco
34	36	52.9	128	6	ABU01997	Abu01997 S. pneumo
35	36	52.9	155	4	AAAB19895	Aab19895 Neisseria
36	36	52.9	174	2	AAW04894	Aaw04894 Proteinas
37	36	52.9	174	2	AAW04893	Aaw04893 Proteinas
38	36	52.9	174	2	AAW04891	Aaw04891 Proteinas
39	36	52.9	174	5	ABG91063	Abg91063 Neisseria
40	36	52.9	174	6	ABU79079	Abu79079 N. mening
41	36	52.9	175	2	AAW04892	Aaw04892 Proteinas
42	36	52.9	175	6	ABP77991	Abp77991 N. gonorr
43	36	52.9	295	4	ABBA62469	Abba62469 Drosophil
44	36	52.9	295	4	ABBA67105	Abba67105 Drosophil
45	36	52.9	333	6	ABM69430	Abm69430 Phototrab

ALIGNMENTS

RESULT 1  
AAE10544  
ID AAE10544 standard; peptide; 13 AA.

AC AAE10544;

XX XX

DT 10-DEC-2001 (first entry)

XX XX

DE Llama species antibody VHH CDR3 #13.

XX XX

KW Llama antibody; camelid; anorectic; heavy chain variable domain; VHH;

KW human dietary enzyme inhibitor; medicament; human pancreatic lipase; HPL;

KW food; human gastric lipase; HGL; cosmetic control; body weight;

KW complementarity determining region 3; CDR3.

XX OS Lama sp.

XX XX

PN EP1134231-A1.

XX XX

PD 19-SEP-2001.

XX XX

PF 20-FEB-2001; 2001EP-00200703.

XX XX

PR 14-MAR-2000; 2000EP-00200930.

XX XX

PA (UNIL ) UNILEVER NV.

XX XX

PI (UNIL ) UNILEVER PLC.

XX XX

PI Bezemer S, Van De Burg M, De Haard JJW, Tareilus E;

XX XX

DR WPI; 2001-572718/65.

XX XX

PT New antibody or its fragments for inhibiting human dietary enzymes.

PT useful for cosmetic control of body weight of human beings, comprises

PT heavy chain variable domain derived from immunoglobulin naturally devoid

PT of light chains.

XX XX

PS Claim 4; Page 29; 37pp; English.

XX XX

CC The patent discloses antibodies or their fragments comprising a heavy

CC chain variable domain (VHH) derived from an immunoglobulin naturally

CC devoid of light chains specific for inhibiting human dietary enzymes. The

CC antibodies of the invention are useful for the preparation of medicaments

CC or food for inhibiting the activity of one or more human dietary enzymes

CC especially human pancreatic lipase (HPL) or human gastric lipase (HGL)

CC which are useful for the cosmetic control of body weight of human beings.

CC The present peptide sequence is a complementarity determining region 3

CC (CDR3) of llama species (camelid) antibody VHH region

```

XX SQ Sequence 13 AA;
Query Match 100.0%; Score 68; DB 4; Length 13;
Best Local Similarity 100.0%; Pred. No. 1.2e-05;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QVRVFSSDYTNV 13
Db 1 QVRVFSSDYTNV 13

RESULT 2
AAE10557
ID AAE10557 standard; peptide; 130 AA.
XX AC AAE10557;
XX DT 10-DEC-2001 (first entry)
XX DE HPL inhibiting VHH fragment, HPL #19 from llama species.
XX KW Llama antibody; camelid; anorectic; heavy chain variable domain; VHH;
XX KW human dietary enzyme inhibitor; medicament; human pancreatic lipase; HPL;
XX KW food; human gastric lipase; HGL; cosmetic control; body weight.
XX OS Lama sp.
XX FH Key
XX FT Region
XX FT 31..35
XX FT /label= CDR1
XX FT /note= "Complementarity determining region 1"
XX FT Region
XX FT 50..64
XX FT /label= CDR2
XX FT /note= "Complementarity determining region 2"
XX FT Region
XX FT 98..110
XX FT /label= CDR3
XX FT /note= "Complementarity determining region 3"
XX EP1134231-A1.
XX PD 19-SEP-2001.
XX PF 20-FEB-2001; 2001EP-00200703.
XX PR 14-MAR-2000; 2000EP-00200930.
XX PA (UNIL ) UNILEVER NV.
XX PA (UNIL ) UNILEVER PLC.
XX PI Bezemer S, Van De Burg M, De Haard JWM, Tareilus B;
XX WPI; 2001-572718/65.
XX DR
XX PT New antibody or its fragments for inhibiting human dietary enzymes,
XX PT useful for cosmetic control of body weight of human beings, comprises
XX PT heavy chain variable domain derived from immunoglobulin naturally devoid
XX PT of light chains.
XX PS Example 2; Page 10; 37pp; English.
XX CC The patent discloses antibodies or their fragments comprising a heavy
XX CC chain variable domain (VHH) derived from an immunoglobulin naturally
XX CC devoid of light chains specific for inhibiting human dietary enzymes. The
XX CC antibodies of the invention are useful for the preparation of medicaments
XX CC or food for inhibiting the activity of one or more human dietary enzymes
XX CC especially human pancreatic lipase (HPL) or human gastric lipase (HGL)
XX CC which are useful for the cosmetic control of body weight of human beings.
XX CC The present peptide sequence is HPL inhibiting VHH fragment, HPL #19 from
XX CC llama (camelid) species
XX SQ Sequence 130 AA;
Query Match 63.2%; Score 43; DB 5; Length 293;
Best Local Similarity 70.0%; Pred. No. 16;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Query Match 100.0%; Score 68; DB 4; Length 130;
Best Local Similarity 100.0%; Pred. No. 0.00017;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QVRVFSSDYTNV 13
Db 98 QVRVFSSDYTNV 110

RESULT 3
ABP26437
ID ABP26437 standard; protein; 293 AA.
XX AC ABP26437;
XX DT 02-JUL-2002 (first entry)
XX DE Streptococcus polypeptide SEQ ID NO 2050.
XX KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
XX KW group A streptococcus; Streptococcus pyogenes; antibacterial;
XX KW antiinflammatory; infection; vaccine; meningitis; gene therapy.
XX OS Streptococcus agalactiae.
XX PN WO200234771-A2.
XX PD 02-MAY-2002.
XX PF 29-OCT-2001; 2001WO-GB004789.
XX PR 27-OCT-2000; 2000GB-00026333.
XX PR 24-NOV-2000; 2000GB-00028727.
XX PR 07-MAR-2001; 2001GB-00005640.
XX PA (CHIR-) CHIRON SPA.
XX PA (GENO-) INST GENOMIC RES.
XX PI Telford J, Massignani V, Margarit Y Rosi, Grandi G, Fraser C;
XX PI Tettelin H;
XX DR WPI; 2002-352536/38.
XX DR N-PSDB; AEN67068.
XX PT New Streptococcus protein for the treatment or prevention of infection or
XX PT disease caused by Streptococcus bacteria, such as meningitis, and for
XX PT detecting a compound that binds to the protein.
XX PS Claim 1; Page 3356; 4525pp; English.
XX CC The invention relates to a protein (ABP25413-ABP30895) from group B
XX CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
XX CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
XX CC the specification. The proteins have antibacterial and antiinflammatory
XX CC activity. (I), nucleic acids encoding (I), ABN6044-ABN71526 and
XX CC antibodies that bind (I) are used in the manufacture of medicaments for
XX CC the treatment or prevention of infection or disease caused by
XX CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
XX CC Nucleic acids encoding (I) are used to detect Streptococcus in a
XX CC biological sample. (I) is used to determine whether a compound binds to
XX CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be
XX CC used as a vaccine or diagnostic composition. The disease caused by
XX CC Streptococcus that is prevented or treated may be meningitis. Nucleic
XX CC acid encoding (I) may be used to recombinantly produce (I) and may be
XX CC used in gene therapy. Antibodies to (I) are used for affinity
XX CC chromatography, immunoassays, and distinguishing/identifying
XX CC Streptococcus proteins
XX SQ Sequence 293 AA;
Query Match 63.2%; Score 43; DB 5; Length 293;
Best Local Similarity 70.0%; Pred. No. 16;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

```

QY 4 VRFSSDYNTY 13  
 Db 17 VRYSGDYTNF 26

RESULT 4  
 ABP30089  
 ID ABP30089 standard; protein; 513 AA.

AC ABP30089;  
 DT 02-JUL-2002 (first entry)  
 DE Streptococcus polypeptide SEQ ID NO 9354.

KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;  
 KW group A streptococcus; Streptococcus pyogenes; antibacterial;  
 KW antiinflammatory; infection; vaccine; meningitis; gene therapy.  
 XX Streptococcus agalactiae.

OS WO200234771-A2.

PN 02-MAY-2002.

XX 29-OCT-2001; 2001WO-GB004789.

XX 27-OCT-2000; 2000GB-00026333.

PR 24-NOV-2000; 2000GB-00028727.

PR 07-MAR-2001; 2001GB-00005640.

XX (CHIR-) CHIRON SPA.

PA (GENO-) INST GENOMIC RES.

XX Telford J, Maignani V, Margarit Y Rosi, Grandi G, Fraser C;

PI Tettelin H;

PI WPI; 2002-352536/38.

DR N-PSDB; ABN70720.

XX New Streptococcus protein for the treatment or prevention of infection or

PT disease caused by Streptococcus bacteria, such as meningitis, and for

PT detecting a compound that binds to the protein.

XX Claim 1; Page 4060; 4525pp; English.

XX The invention relates to a protein (ABP25413-ABP30895) from group B

CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS

CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in

CC the specification. The proteins have antibacterial and antiinflammatory

CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and

CC antibodies that bind (I) are used in the manufacture of medicaments for

CC the treatment or prevention of infection or disease caused by

CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.

CC Nucleic acids encoding (I) are used to detect Streptococcus in a

CC biological sample. (I) is used to determine whether a compound binds to

CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be

CC used as a vaccine or diagnostic composition. The disease caused by

CC Streptococcus that is prevented or treated may be meningitis. Nucleic

CC acid encoding (I) may be used to recombinantly produce (I) and may be

CC used in gene therapy. Antibodies to (I) are used for affinity

CC chromatography, immunoassays, and distinguishing/identifying

CC Streptococcus proteins

XX Sequence 513 AA;

QY Query Match 63.2%; Score 43; DB 5; Length 513;

Best Local Similarity 70.0%; Pred. No. 30;

Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 237 VRYSGDYTNF 246.  
 RESULT 5  
 AAO17579  
 ID AAO17579 standard; protein; 208 AA.  
 XX  
 AC AAO17579;  
 DT 19-JUL-2002 (first entry)  
 DE M catarrhalis MCA101792 protein SEQ ID NO: 38.  
 KW Moraxella; vaccine; respiratory tract infection; antiinflammatory;  
 KW auditory; antibacterial; otitis media; sinusitis; pneumonia.  
 XX Moraxella catarrhalis.  
 OS WO200218595-A2.  
 PN 07-MAR-2002.  
 PD  
 XX 28-AUG-2001; 2001WO-CA001221.  
 XX 28-AUG-2000; 2000US-0228294P.  
 PR 28-AUG-2000; 2000US-0228295P.  
 PR 28-AUG-2000; 2000US-0228296P.  
 PR 29-AUG-2000; 2000US-0228438P.  
 PR 29-AUG-2000; 2000US-0228439P.  
 PR 29-AUG-2000; 2000US-0228440P.  
 PR 29-AUG-2000; 2000US-0228441P.  
 PR 29-AUG-2000; 2000US-0228442P.  
 PR 29-AUG-2000; 2000US-0228443P.  
 PR 29-AUG-2000; 2000US-0228511P.  
 PR 29-AUG-2000; 2000US-0228512P.  
 PR 29-AUG-2000; 2000US-0228742P.  
 PR 29-AUG-2000; 2000US-0228773P.  
 PR 01-SEP-2000; 2000US-0229465P.  
 PR 01-SEP-2000; 2000US-0229474P.  
 PR 01-SEP-2000; 2000US-0229475P.  
 PR 01-SEP-2000; 2000US-0229478P.  
 PR 01-SEP-2000; 2000US-0229740P.  
 PR 05-SEP-2000; 2000US-0229803P.  
 PR 05-SEP-2000; 2000US-0229804P.  
 PR 05-SEP-2000; 2000US-0229805P.  
 PR 05-SEP-2000; 2000US-0229806P.  
 PR 05-SEP-2000; 2000US-0229809P.  
 PR 05-SEP-2000; 2000US-0229811P.  
 PR 06-SEP-2000; 2000US-0230214P.  
 PR 06-SEP-2000; 2000US-0230250P.  
 PR 06-SEP-2000; 2000US-0230252P.  
 XX (AVET) AVENTIS PASTEUR LTD.  
 XX Loesmore S, Wang J, Bradley B, Ochs M, Yang Y;  
 WPI; 2002-404555/43.  
 DR N-PSDB; AAU46511.  
 XX Moraxella polypeptide and polynucleotides useful as vaccine for  
 PT immunizing a host e.g. humans against disease e.g. otitis media,  
 PT pneumonia, caused by infection of the bacteria.  
 XX Claim 28; Fig 37; 277pp; English.  
 XX The present invention provides the protein and coding sequences of  
 CC proteins from Moraxella catarrhalis. These can be used to produce  
 CC vaccines which protect against M. catarrhalis infection, which can cause  
 CC otitis media, respiratory infection, sinusitis, and pneumonia. The  
 CC present sequence is a protein of the invention  
 XX Sequence 208 AA;

Query Match 61.8%; Score 42; DB 5; Length 208;  
 Best Local Similarity 60.0%; Pred. No. 16;  
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 4 VRFSDYTNV 13  
 :||:||||:  
 Db 64 IRFQADYTNF 73

## RESULT 6

ABP33318  
 ID ABP33318 standard; protein; 117 AA.

XX AC ABP33318;

XX DT 09-JUL-2002 (first entry)

XX DE Human ORF2291 protein, SEQ ID NO:4582.

XX KW Human; ORF; open reading frame; ORFX; drug screening; diagnosis;  
 KW disease monitoring; cytokine; cell proliferation; cell differentiation;  
 KW immune modulation; haematopoiesis regulation; tissue growth;  
 KW angiogenesis; activin; inhibit; chemotactic; chemokinetic; haemostatic;  
 KW thrombolytic; tumour inhibition; bodily characteristic; fertility;  
 KW behaviour; cancer; proliferative disorder; neurological disorder;  
 KW cardiovascular disease; immune system disorder; organ transplantation;  
 KW tissue growth disorder; tissue regeneration disorder; diabetes mellitus;  
 KW hypothyroidism; cholesterol ester storage disease; infection; vulnery;  
 KW vasotropic; antipsoriatic; antidiabetic; cytostatic; neutropic;  
 KW neuroprotective; antiatherosclerotic; anticoagulant; thrombolytic;  
 KW cardiant; hypotensive; antithyroid; antiinflammatory; immunomodulator;  
 KW dermatological; analgesic; virucide; antibacterial; fungicide.

XX OS Homo sapiens.

XX PN WO200190366-A2.

XX PD 29-NOV-2001.

XX PF 24-MAY-2001; 2001WO-US017076.

XX PR 24-MAY-2000; 2000US-0206690P.

XX PA (CURA-) CURAGEN CORP.

XX PI Leach MD, Shimkets RA;

XX DR WPI; 2002-106200/14.

XX DR N-PSDB; ABN77344.

XX PT Novel human polypeptides and polynucleotides useful for diagnosing,  
 PT preventing and treating cardiovascular disease, neurodegenerative,  
 PT hyperproliferative disorders and disorders related to organ  
 PT transplantation.

XX PS Claim 10; Page 1400; 2508pp; English.

XX CC Sequences ABP31028-ABP35561 represent 4534 novel human proteins  
 CC designated ORF (open reading frame) 1-4534, and sequences ABN75054-  
 CC ABN79587 represent cDNAs encoding them. The invention also encompasses  
 CC polypeptides at least 80% identical to the ORF1-ORF4534 (collectively  
 CC referred to as ORFX) proteins, polynucleotides at least 85% identical to  
 CC the ORFX nucleic acid sequences, vectors and host cells comprising ORFX  
 CC polynucleotides, the recombinant production of ORFX proteins, antibodies  
 CC specific for ORFX proteins, methods of detecting ORFX polynucleotides and  
 CC polypeptides, methods of screening for modulators of ORFX expression or  
 CC activity, and methods of screening individuals for a predisposition to an  
 CC ORFX-associated disorder. The ORFX proteins of the invention have a wide  
 CC range of biological activities, such as cytokine, cell proliferation,  
 CC cell differentiation, immune modulation, haematopoiesis regulation,  
 CC tissue growth, angiogenesis, activin or inhibit activity, chemotactic/  
 CC chemokinetic activity, haemostatic activity, thrombolytic activity,  
 CC receptor/ligand, antiinflammatory activity, tumour inhibition activity,

CC and antinfective activity, and may also be involved in the determination  
 CC of bodily characteristics, fertility and behaviour. ORFX proteins,  
 CC nucleic acids and antibodies may be used in the treatment of cancers,  
 CC other proliferative disorders such as psoriasis and benign tumours,  
 CC neurological disorders such as epilepsy and Alzheimer's disease,  
 CC cardiovascular diseases, immune system disorders, disorders related to  
 CC organ transplantation, disorders of tissue growth and regeneration,  
 CC diseases such as diabetes mellitus, hypothyroidism, and cholesterol  
 CC storage disease, and infectious diseases caused by viral, bacterial,  
 CC fungal and other pathogens. ORFX nucleic acids may also be used as a  
 CC source of primers and probes in the detection of ORFX genomic sequences  
 CC or transcripts, in the identification and cloning of homologous  
 CC sequences, in genetic diagnosis, and in forensic biology. The ORFX  
 CC nucleic acids may additionally be used to produce transgenic animals  
 CC which may be useful for studying the function and/or activity of ORFX  
 CC protein, and in drug screening. The ORFX proteins may also be used as  
 CC immunogens to generate specific antibodies, which are useful in the  
 CC diagnosis, treatment and monitoring of ORFX-associated diseases

XX SQ Sequence 117 AA;

Query Match 60.3%; Score 41; DB 5; Length 117;

Best Local Similarity 88.9%; Pred. No. 13;

Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 VVRFRSSDY 10

Db 3 VVRFRSGDY 11  
 |||||

## RESULT 7

AAV81486

ID AAV81486 standard; protein; 476 AA.

XX AC AAV81486;

XX DT 03-JUL-2000 (first entry)

XX DE Pseudomonas putida aldehyde dehydrogenase.

XX KW Aldehyde dehydrogenase; NAD+-dependent; oxidation; 2-oxoaldehyde;

XX KW hydroxypyruvaldehyde; glyceraldehyde; methylglyoxal; hydroxypyruvic acid.

XX OS Pseudomonas putida.

XX PN JP2000060563-A.

XX PD 29-FEB-2000.

XX PF 24-AUG-1998; 98JP-00236691.

XX PR 24-AUG-1998; 98JP-00236691.

XX PA (CHIK-) ZH CHIKYU KANKYO SANGYO GIJITSU KENKYU.

XX DR WPI; 2000-249678/22.

XX DR N-PSDB; AAA07084.

XX PT New aldehyde dehydrogenase gene - for enzymatic synthesis of serine or  
 PT cysteine from glycerol.

XX PS Claim 2; Page 8-9; 11pp; Japanese.

XX CC This sequence represents a novel aldehyde dehydrogenase from Pseudomonas  
 CC putida. This catalyses reactions such as the NAD+-dependent oxidation of  
 CC hydroxypyruvaldehyde to hydroxypyruvic acid and reacts specifically with  
 CC 2-oxoaldehydes such as hydroxypyruvaldehyde, glyceraldehyde and  
 CC methylglyoxal. The protein was initially purified from a culture of P.  
 CC putida, and subjected to tryptic cleavage. The fragments thus obtained  
 CC were sequenced, and the sequence information used to design primers  
 CC AAA07085-A07086. These primers were used to amplify a aldehyde  
 CC dehydrogenase PCR product. The fragment was used in the detection of a  
 CC full-length DNA encoding aldehyde dehydrogenase in a genomic library.



CC This was amplified using primers AAA07087-A07088, and inserted into a  
 CC vector for transformation of cells. Recombinantly produced Pseudomonas  
 CC putida aldehyde dehydrogenase may be used for the synthesis of serine or  
 CC cysteine from glycerol  
 CC  
 SQ Sequence 476 AA;

Query Match 60.3%; Score 41; DB 3; Length 476;  
 Best Local Similarity 53.8%; Pred. No. 64;  
 Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;  
 QY 1 QVRVRFSSDYTY 13  
 |||:|:|:|:  
 Db 99 QVEVNFTADYLDY 111

RESULT 8  
 ABU40154  
 ID ABU40154 standard; protein; 748 AA.  
 XX  
 AC ABU40154;  
 XX  
 XX 19-JUN-2003 (first entry)  
 XX  
 DE Protein encoded by Prokaryotic essential gene #25681.  
 XX  
 KW Antisense; prokaryotic essential gene; cell proliferation; drug design.  
 XX  
 OS Pseudomonas putida.  
 XX  
 XX WO20027183-A2.  
 PN  
 XX  
 XX 03-OCT-2002.  
 PD  
 XX  
 XX 21-MAR-2002; 2002WO-US009107.  
 PF  
 XX  
 XX 21-MAR-2001; 2001US-00815242.  
 PR  
 XX 06-SEP-2001; 2001US-00948993.  
 PR  
 XX 25-OCT-2001; 2001US-0342923P.  
 PR  
 XX 08-FEB-2002; 2002US-00072851.  
 PR  
 XX 06-MAR-2002; 2002US-0362699P.  
 PR  
 XX (ELIT-) ELITRA PHARM INC.  
 PA

XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
 XX Wall D, Trawick JD, Carr GU, Yamamoto R, Forsyth RA, Xu HH;  
 XX WPI; 2003-029926/02.  
 XX N-PSDB; ACA44024.  
 XX  
 XX New antisense nucleic acids, useful for identifying proteins or screening  
 XX for homologous nucleic acids required for cellular proliferation to  
 XX isolate candidate molecules for rational drug discovery programs.  
 XX  
 XX Claim 25; SEQ ID NO 68078; 1766pp; English.

XX The invention relates to an isolated nucleic acid comprising any one of  
 XX the 6213 antisense sequences given in the specification where expression  
 XX of the nucleic acid inhibits proliferation of a cell. Also included are:  
 XX (1) a vector comprising a promoter operably linked to the nucleic acid  
 XX encoding a polypeptide whose expression is inhibited by the antisense  
 XX nucleic acid; (2) a host cell containing the vector; (3) an isolated  
 XX polypeptide or its fragment whose expression is inhibited by the  
 XX antisense nucleic acid; (4) an antibody capable of specifically binding  
 XX the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
 XX proliferation or the activity of a gene in an operon required for  
 XX the gene product or that has an activity against a biological pathway  
 XX required for proliferation, or that inhibits cellular proliferation; (8)  
 XX identifying a gene required for cellular proliferation or the biological  
 XX pathway in which a proliferation-required gene or its gene product lies  
 XX or a gene on which the test compound that inhibits proliferation of an  
 XX organism acts; (9) manufacturing an antibiotic; (10) profiling a

CC compound's activity; (11) a culture comprising strains in which the gene  
 CC product is overexpressed or underexpressed; (12) determining the extent  
 CC to which each of the strains is present in a culture or collection of  
 CC strains; or (13) identifying the target of a compound that inhibits the  
 CC proliferation of an organism. The antisense nucleic acids are useful for  
 CC identifying proteins or screening for homologous nucleic acids required  
 CC for cellular proliferation to isolate candidate molecules for rational  
 CC drug discovery programs, or for screening homologous nucleic acids  
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
 CC the target prokaryotic essential genes. Note: The sequence data for this  
 CC patent did not form part of the printed specification, but was obtained  
 CC in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 748 AA;

Query Match 60.3%; Score 41; DB 6; Length 748;  
 Best Local Similarity 88.9%; Pred. No. 11e+02;  
 Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 VVRVRFSSDY 10  
 |||||:  
 Db 102 VVRVRFSGDY 110

RESULT 9  
 AAW00456  
 ID AAW00456 standard; protein; 615 AA.  
 XX  
 AC AAW00456;

XX  
 XX 16-OCT-2003 (revised)  
 DT 25-SEP-1996 (first entry)  
 DT  
 XX  
 DE 1-Fructan:fructan fructosyltransferase.  
 XX  
 XX 1-fructan:fructan fructosyltransferase; SST; fructan; sweetener;  
 KW surfactant; transgenic plant; Jerusalem artichoke.  
 XX  
 XX Helianthus tuberosus; cv. Colombia.  
 OS  
 XX WO9621023-A1.  
 PN  
 XX 11-JUL-1996.  
 PD  
 XX 08-JAN-1996; 96WO-NL000012.  
 PF  
 XX 06-JAN-1995; 95EP-00200015.  
 PR 27-MAR-1995; 95EP-00200762.  
 PR  
 XX (CPRO-) CPRO-DLO CENT PLANTENVEREDELINGS REPROD.  
 PA  
 XX Van Tunen AJ, Van Der Meer IM, Koops AJ;  
 PI  
 XX WPI; 1996-333998/33.  
 XX N-PSDB; AAT31025.

XX DNA sequences encoding fructan-producing enzymes - used to transform  
 XX plants, for the production of low calorie sweeteners etc.  
 XX  
 XX Example 1; Fig 4b; Slipp; English.  
 XX  
 XX Jerusalem artichoke 1-fructan:fructan fructosyltransferase (FFT)  
 XX catalyses a polymerisation reaction involving the transfer of fructosyl  
 XX units between trisaccharides, tetrasaccharides and larger fructose  
 XX polymers. Its amino acid sequence was deduced from a cDNA clone  
 XX (AA31025) isolated from a Jerusalem artichoke tuber cDNA library. The  
 XX enzyme can be expressed in transgenic plants to modify the fructan  
 XX profile. The fructans are useful as low-calorie sweeteners or as the  
 XX hydrophilic moiety in biosurfactants. (Updated on 16-OCT-2003 to  
 XX standardise OS field)

SQ Sequence 615 AA;  
 Query Match 58.8%; Score 40; DB 2; Length 615;  
 Best Local Similarity 58.3%; Pred. No. 1.3e+02;  
 Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 2 VVRFSSDYTN 13  
 DB 252 VLVYTTDTNY 263

RESULT 10  
 AAB51367  
 ID AAB51367 standard; protein; 3132 AA.  
 AC AAB51367;  
 XX 10-APR-2001 (first entry)  
 DT  
 DE Japanese yam mosaic virus JYMV-M protein sequence SEQ ID NO:1.  
 XX  
 KW Japanese yam mosaic virus; JYMV; Dioscorea; resistance; JYMV-M;  
 KW Dioscoreaceae mosaic virus.  
 XX  
 OS Japanese yam mosaic virus.  
 XX JP2000300270-A.  
 PN  
 XX 31-OCT-2000.  
 PD  
 XX 21-APR-1999; 99JP-00113671.  
 PF  
 XX 21-APR-1999; 99JP-00113671.  
 PR  
 XX (AICH-) AICHI KEN PREFECTURE.  
 PA  
 XX WPI; 2001-141353/15.  
 DR N-PSDB; AAF26998.  
 XX

Weakly toxic strain of Dioscoreaceae mosaic virus used in the genetic diagnosis of the mosaic virus infection and for the management of crops resistant to the virus.

Claim 1; Page 6-20; 22pp; Japanese.

CC The present sequence represents the Japanese yam mosaic virus (JYMV) JYMV -M protein sequence. JYMV is a weakly toxic strain of Dioscoreaceae mosaic virus. The present invention also describes: (1) a weakly toxic strain of Dioscoreaceae mosaic virus having a nucleic acid sequence which shows interference against the infection of a strongly toxic strain of Dioscoreaceae mosaic virus; (2) a Dioscoreaceae mosaic virus-resistant crop which is a plant of Dioscorea japonica, Dioscorea opposita or Dioscorea alata and is infected artificially by a weakly toxic strain of Dioscoreaceae mosaic virus; (3) a method for the cultivation of a Dioscoreaceae mosaic virus-resistant crop in which the weakly toxic strain of Dioscoreaceae mosaic virus is infected to the tubercle or the bulb of a plant of Dioscorea japonica, Dioscorea opposita or Dioscorea alata and the tubercle or bulb is grown; and (4) a method for genetic diagnosis of a Dioscoreaceae mosaic virus in which a DNA fragment amplified by using reverse transcriptase-polymerase chain reaction (RT-PCR) primers atagctacacacccctac (AAF2699) and caccctacatggwagayg (AAF32201) is treated with a restriction enzyme to distinguish the weakly toxic strain of Dioscoreaceae mosaic virus from the strongly toxic strain of Dioscoreaceae mosaic virus

SQ Sequence 3132 AA;  
 Query Match 58.8%; Score 40; DB 4; Length 3132;  
 Best Local Similarity 70.0%; Pred. No. 8.3e+02;  
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 2 VVRFSSDYTN 11  
 DB 252 VLVYTTDTNY 263

RESULT 11  
 AAG73797  
 ID AAG73797 standard; protein; 176 AA.  
 AC AAG73797;  
 XX 03-SEP-2001 (first entry)  
 DT  
 DE Human colon cancer antigen protein SEQ ID NO:4561.  
 XX  
 KW Human; colon cancer; colon cancer antigen; diagnosis; detection; colorectal carcinoma.  
 KW  
 XX Homo sapiens.  
 OS  
 XX WO200122920-A2.  
 PN  
 XX 05-APR-2001.  
 PD  
 PF 28-SEP-2000; 2000WO-US026524.  
 XX  
 PR 29-SEP-1999; 99US-0157137P.  
 PR 03-NOV-1999; 99US-0163280P.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 XX Ruben SM, Barash SC, Birse CE, Rosen CA;  
 PI  
 XX WPI; 2001-235357/24.  
 DR N-PSDB; AAH33228.  
 XX

Nucleic acids encoding 4277 human colon cancer-associated polypeptides, useful for preventing, diagnosing and/or treating colorectal cancers.

Claim 11; Page 6360-6361; 9803pp; English.

CC AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon cancer-associated nucleic acid molecules (N) and proteins (P), where the proteins are collectively known as colon cancer antigens. The colon cancer antigens have cytostatic activity and can be used in gene therapy and vaccine production. N and P may be used in the prevention, diagnosis and treatment of diseases associated with inappropriately P expression. For example, N and P may be used to treat disorders associated with decreased expression by rectifying mutations or deletions in a patient's genome that affect the activity of P by expressing inactive proteins or to supplement the patients' own production of P. Additionally, N may be used to produce the colon cancer-associated Ps, by inserting the nucleic acids into a host cell and culturing the cell to express the proteins. N and P can be used in the prevention, diagnosis and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204 and AAG77789 represent sequences used in the exemplification of the present invention. N.B. Pages 666 to 682 and page 7053 of the sequence listing were missing at time of publication, meaning no sequences are present for SEQ ID NO:1027 to 1052, 7921 and 7922

XX Sequence 176 AA;  
 Query Match 57.4%; Score 39; DB 4; Length 176;  
 Best Local Similarity 53.8%; Pred. No. 47;  
 Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 1 QVRFSSDYTN 13  
 DB 50 QVRFSSDYTN 62

RESULT 12  
 ABB47642  
 ID ABB47642 standard; protein; 407 AA.  
 XX

AC AB47642;  
 XX 05-FEB-2002 (first entry)  
 DE Listeria monocytogenes protein #346.  
 XX Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;  
 KW vitamin B12; bacterial infection; disease.  
 XX Listeria monocytogenes.  
 OS WO200177335-A2.  
 XX 18-OCT-2001.  
 XX 11-APR-2001; 2001WO-FR001118.  
 PF 11-APR-2000; 2000FR-00004629.  
 PR (INSP ) INST PASTEUR.  
 PA Buchrieser C, Frangeul L, Couve E, Rusniok C, Fsihi H, Dehoux P;  
 PI Dussurgeat O, Chetoui F, Nedjari H, Glaser P, Kunst F, Cossart P;  
 PI Daniels J, Giesbel W, Kref J, Kuhn M, Ng E, Vazquez-Boland JA;  
 PI Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A;  
 PI Chakraborty T, Domann E, Hain T, Berche P, Charbit A, Durand L;  
 PI Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;  
 PI Maduenio E, De Pablos B, Wehland J, Kaerst U, Entian K, Hauf J;  
 PI Rose M, Voss H;  
 XX WPI; 2002-010914/01.  
 XX Genomic sequence for Listeria monocytogenes, useful e.g. for treatment  
 PT and prevention of Listeria and related bacterial infections, and related  
 PT polypeptides.  
 XX Claim 6; SEQ ID NO 347; 192bp; French.  
 XX The present invention relates to the genome sequence of Listeria  
 CC monocytogenes EGD-e (see ABA03041). The genome sequence and fragments of  
 CC it are useful for selecting probes and primers for detecting genes in L.  
 CC polycytophisms and related organisms, and for studying genetic  
 CC polymorphisms and other genomes. The present sequence is a protein  
 CC encoded by the genome sequence of the present invention. Proteins  
 CC expressed from the genome sequence are useful for raising specific  
 CC antibodies, identification of L. monocytogenes and related organisms, and  
 CC for biosynthesis and biodegradation, especially biosynthesis of vitamin  
 CC B12. The genome sequence and proteins encoded by it are also useful for  
 CC selecting compounds that regulate gene expression and cell replication  
 CC and modulate L. monocytogenes-related diseases. In addition, the genome  
 CC sequence and proteins encoded by it are useful in pharmaceutical and  
 CC vaccine compositions for the treatment or prevention of infections by L.  
 CC monocytogenes and related organisms. Note: The sequence data for this  
 CC patent did not form part of the printed specification, but was obtained  
 CC in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX Sequence 407 AA;  
 SQ Query Match 57.4%; Score 39; DB 5; Length 407;  
 Best Local Similarity 63.6%; Pred. No. 1.2e+02;  
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
 Oy 1 QVRVRFSSDYT 11  
 |||||  
 Db 330 QVRKMASDYT 340  
 RESULT 13  
 ABU32421  
 ID ABU32421 standard; protein; 407 AA.  
 XX  
 AC ABU32421;

XX 19-JUN-2003 (first entry)  
 XX Protein encoded by Prokaryotic essential gene #17948.  
 DE Antisense; prokaryotic essential gene; cell proliferation; drug design.  
 XX Listeria monocytogenes.  
 OS WO200277183-A2.  
 XX 03-OCT-2002.  
 XX 21-MAR-2002; 2002WO-US009107.  
 XX 21-MAR-2001; 2001US-00815242.  
 PR 06-SEP-2001; 2001US-00948993.  
 PR 25-OCT-2001; 2001US-0342923P.  
 PR 08-FEB-2002; 2002US-00072851.  
 PR 06-MAR-2002; 2002US-0362699P.  
 XX (ELIT-) ELITRA PHARM INC.  
 PA Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
 PI Wall D, Trawick JD, Carr GU, Yamamoto R, Forsyth RA, Xu HH;  
 XX WPI; 2003-029926/02.  
 DR N-PSDB; ACA36291.  
 XX New antisense nucleic acids, useful for identifying proteins or screening  
 PT for homologous nucleic acids required for cellular proliferation to  
 PT isolate candidate molecules for rational drug discovery programs.  
 XX Claim 25; SEQ ID NO 60345; 1766pp; English.  
 XX The invention relates to an isolated nucleic acid comprising any one of  
 CC the 6213 antisense sequences given in the specification where expression  
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
 CC (1) a vector comprising a promoter operably linked to the nucleic acid  
 CC encoding a polypeptide whose expression is inhibited by the antisense  
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
 CC polypeptide or its fragment whose expression is inhibited by the  
 CC antisense nucleic acid; (4) an antibody capable of specifically binding  
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
 CC proliferation or the activity of a gene in an operon required for  
 CC proliferation; (7) identifying a compound that influences the activity of  
 CC the gene product or that has an activity against a biological pathway; (8)  
 CC required for proliferation, or that inhibits cellular proliferation; (8)  
 CC identifying a gene required for cellular proliferation or the biological  
 CC pathway in which a proliferation-required gene or its gene product lies  
 CC or a gene on which the test compound that inhibits proliferation of an  
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
 CC compound's activity; (11) a culture comprising strains in which the gene  
 CC product is overexpressed or underexpressed; (12) determining the extent  
 CC to which each of the strains is present in a culture or collection of  
 CC strains; or (13) identifying the target of a compound that inhibits the  
 CC proliferation of an organism. The antisense nucleic acids are useful for  
 CC identifying proteins or screening for homologous nucleic acids required  
 CC for cellular proliferation to isolate candidate molecules for rational  
 CC drug discovery programs, or for screening homologous nucleic acids  
 CC required for proliferation in cells other than S. aureus, S. typhimurium,  
 CC K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of  
 CC the target prokaryotic essential genes. Note: The sequence data for this  
 CC patent did not form part of the printed specification, but was obtained  
 CC in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX Sequence 407 AA;  
 SQ Query Match 57.4%; Score 39; DB 6; Length 407;  
 Best Local Similarity 63.6%; Pred. No. 1.2e+02;  
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 QVRVFPSSDYT 11  
 Db 330 QVRXNASDYT 340  
 RESULT 14  
 ID ABB62215 standard; protein; 469 AA.  
 AC ABB62215;  
 XX 26-MAR-2002 (first entry)  
 XX Drosophila melanogaster polypeptide SEQ ID NO 13437.  
 DE Drosophila; developmental biology; cell signalling; insecticide;  
 XX pharmaceutical.  
 KW Drosophila melanogaster.  
 XX WO200171042-A2.  
 XX 27-SEP-2001.  
 PD 23-MAR-2001; 2001WO-US009231.  
 XX 23-MAR-2000; 2000US-0191637P.  
 PR 11-JUL-2000; 2000US-00614150.  
 XX (PEKE) PE CORP NY.  
 PA Venter JC, Adams M, Li PWD, Myers EW;  
 PI WPI; 2001-656860/75.  
 DR N-PSDB; ABL06318.  
 XX New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signaling and cell-cell  
 PT interactions.  
 XX Disclosure; SEQ ID NO 13437; 21pp + Sequence Listing; English.  
 PS The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 CC sequences (ABL01840-ABL16175), and the encoded proteins (ABBS7737-  
 CC ABB72072). The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
 XX Sequence 469 AA;  
 SQ  
 Query Match 57.4%; Score 39; DB 4; Length 469;  
 Best Local Similarity 50.0%; Pred. No. 1.5e+02;  
 Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;  
 QY 2 VVRVFPSSDYTNY 13  
 Db 81 LRISFREYTYNY 92  
 RESULT 15  
 ID ABB16347  
 AC ABB16347 standard; protein; 849 AA.  
 XX ABB16347;  
 XX 18-FEB-2002 (first entry)  
 DE Novel human diagnostic protein #16338.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder.  
 XX Homo sapiens.  
 XX WO200175067-A2.  
 XX 11-OCT-2001.  
 XX 30-MAR-2001; 2001WO-US008631.  
 XX 31-MAR-2000; 2000US-00540217.  
 PR 23-AUG-2000; 2000US-00649167.  
 XX (HYSE-) HYSEQ INC.  
 XX Drmanac RT, Liu C, Tang YT;  
 PI WPI; 2001-639362/73.  
 DR N-PSDB; AAS90534.  
 XX New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity.  
 XX Claim 20; SEQ ID NO 46706; 103pp; English.  
 XX The invention relates to isolated polynucleotide (I) and polypeptide (II)  
 CC sequences. (I) is useful as hybridisation probes, polymerase chain  
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
 CC and in recombinant production of (II). The polynucleotides are also used  
 CC in diagnostics as expressed sequence tags for identifying expressed  
 CC genes. (I) is useful in gene therapy techniques to restore normal  
 CC activity of (II) or to treat disease states involving (II). (II) is  
 CC useful for generating antibodies against it, detecting or quantitating a  
 CC polypeptide in tissue, as molecular weight markers and as a food  
 CC supplement. (II) and its binding partners are useful in medical imaging  
 CC of sites expressing (II). (I) and (II) are useful for treating disorders  
 CC involving aberrant protein expression or biological activity. The  
 CC polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABB00010-ABG30377 represent novel human diagnostic  
 CC amino acid sequences of the invention. Note: the sequence data for this  
 CC patent did not appear in the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX Sequence 849 AA;  
 SQ  
 Query Match 57.4%; Score 39; DB 4; Length 849;  
 Best Local Similarity 53.8%; Pred. No. 2.9e+02;  
 Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;  
 QY 1 QVRVFPSSDYTNY 13  
 Db 159 QVKIRFGSDPATY 171  
 Search completed: October 5, 2004, 07:59:43  
 Job time : 14.8904 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2004, 07:36:16; Search time 2.10648 Seconds  
(without alignments)  
593.639 Million cell updates/sec

Title: US-09-805-290A-12

Perfect score: 68

Sequence: 1 LIRRXFTSEYNE 13

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 283366 seqs, 9619126 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database: PIR 78\*

1: PIR1\*

2: PIR2\*

3: PIR3\*

4: PIR4\*

pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	40.5	59.6	413	2 B90162	DNA/pantothenate m
2	40	58.8	383	2 S28422	1,4-alpha-glucan b
3	40	58.8	826	2 T06494	1,4-alpha-glucan b
4	39	57.4	179	2 B90408	hypothetical prote
5	39	57.4	280	2 T18268	hypothetical prote
6	38	55.9	190	2 C89966	hypothetical prote
7	38	55.9	337	2 E90176	conserved hypothet
8	38	55.9	339	2 H97107	uncharacterized pr
9	38	55.9	793	2 C72219	DNA mismatch repa
10	38	55.9	1111	1 A33284	myosin heavy chain
11	37.5	55.1	719	2 S55939	telomerase compone
12	37	54.1	178	2 H70484	probable CDPdiacyl
13	37	54.4	212	2 S56199	hypothetical prote
14	37	54.4	276	2 A82568	polypeptin bios
15	37	54.4	387	2 B90280	galactarate dehydr
16	37	54.4	477	2 A0616	probable ion:amino
17	37	54.4	515	2 T34093	steroid/thyroid/re
18	37	54.4	669	2 S74391	exonuclease ABC c
19	37	54.4	830	2 T07824	1,4-alpha-glucan b
20	37	54.4	854	2 S65045	1,4-alpha-glucan b
21	37	54.4	858	2 B84780	starch branching e
22	37	54.4	861	1 S34730	1,4-alpha-glucan b
23	37	54.4	1215	2 T11660	probable RNA-direc
24	36	52.9	134	2 T29690	hypothetical prote
25	36	52.9	329	2 S67807	hypothetical prote
26	36	52.9	364	2 T29683	hypothetical prote
27	36	52.9	376	2 S51335	probable aryl-alco
28	36	52.9	394	2 A75489	carbamoyl-phosphat
29	36	52.9	400	2 ACl452	hypothetical prote

30 36 52.9 461 2 T08624  
31 36 52.9 466 2 A84535  
32 36 52.9 476 2 T52157  
33 36 52.9 666 2 S34197  
34 36 52.9 874 2 S62486  
35 36 52.9 980 2 T43108  
36 36 52.9 1099 2 C86625  
37 36 52.9 1534 2 A56734  
38 35.5 52.2 391 2 S18666  
39 35 51.5 184 2 S64020  
40 35 51.5 243 2 S13296  
41 35 51.5 295 2 E95200  
42 35 51.5 295 2 D98067  
43 35 51.5 334 2 C85048  
44 35 51.5 352 2 JC7525  
45 35 51.5 410 2 B48585

#### ALIGNMENTS

##### RESULT 1

B90162  
DNA/pantothenate metabolism flavoprotein (dfp) [imported] - Sulfolobus solfataricus  
C:Species: Sulfolobus solfataricus  
C:Date: 24-May-2001 #sequence\_revision 24-May-2001 #text\_change 15-Jun-2001  
C:Accession: B90162  
R:She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Chan-y  
Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, P.  
arrett, R.A.; Ragan, M.A.; Sensesen, C.W.; Van der Oost, J.  
submitted to GenBank, April 2001  
A:Description: Sulfolobus solfataricus complete genome.  
A:Reference number: A99139  
A:Accession: B90162  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-413 <KUR>  
A:Cross-references: CB:AB006641; NID:G13813345; PIDN:AAK40553.1; GSPDB:GN00155  
C:Genetics:  
A:Gene: dfp  
C:Superfamily: pantothenate metabolism flavoprotein dfp

Query Match 59.6%; Score 40.5; DB 2; Length 413;  
Best Local Similarity 64.3%; Pred. No. 23;  
Matches 9; Conservative 2; Mismatches 0; Indels 3; Gaps 1;  
QY 2 IRRK---FTSEYNE 12  
DB 365 VRRKDIGFSSEYNE 378

##### RESULT 2

S28422  
1,4-alpha-glucan branching enzyme (EC 2.4.1.18) - cassava (fragments)  
C:Species: Manihot esculenta (cassava)  
C:Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 07-May-1999  
C:Accession: S28422  
R:Salehuzzaman, S.N.I.M.; Jacobsen, E.; Visser, R.G.F.  
Plant Mol. Biol. 20, 809-819, 1992  
A:Title: Cloning, partial sequencing and expression of a cDNA coding for branching enzymes  
A:Reference number: S28422; MUID:93099233; PMID:1281436  
A:Accession: S28422  
A:Status: not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 1-143;144-298;299-383 <SAL>  
C:Function:  
A:Description: catalyzes the transglycosylation of a terminal segment of a 1,4-alpha-D-g  
A:Pathway: glycogen/starch biosynthesis  
A:Note: final step in biosynthesis of glycogen or amylopectin  
C:Superfamily: 1,4-alpha-glucan branching enzyme  
C:Keywords: glycogen/starch biosynthesis; glycosyltransferase; hexosyltransferase  
Query Match 58.8%; Score 40; DB 2; Length 383;

probable NtrC-type  
Ac-like transposas  
hypothetical prote  
DNA-binding protei  
probable leucyl tr  
lactacin 481/lacto  
subunit B of ATP-d  
ribosome receptor,  
KIN17 protein - mo  
hypothetical prote  
Schwannoma-derived  
fructokinase (impo  
fructokinase (BC 2  
probable transposo  
chondrotin 4-sulf  
transcription fact

```

Best Local Similarity 58.3%; Pred. No. 26;
Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 IRRKFTSEYNEY 13
  ||| : |||
Db 167 INNAFTGDYNEY 178

RESULT 3
T06494
1,4-alpha-glucan branching enzyme (EC 2.4.1.18) II - garden pea (fragment)
N:Alternate names: starch branching enzyme II
C:Species: Pisum sativum (garden pea)
C>Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 21-Jul-2000
C:Accession: T06494
R:Burton, R.A.; Bewley, J.D.; Smith, A.M.; Bhattacharyya, M.K.; Tatge, H.; Ring, S.; Bul-
Plant J. 7, 3-15, 1995
A:Title: Starch branching enzymes belonging to distinct enzyme families are differential
A:Reference number: Z15717; MUID:95201926; PMID:7894509
A:Accession: T06494
A:Molecule type: mRNA
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Cross-references: EMBL:X80010; NID:9510546; PIDN:CAA56320.1; PID:gl345571
C:Genetics:
A:Note: SBETII
C:Superfamily: 1,4-alpha-glucan branching enzyme
C:Keywords: glycogen/starch biosynthesis; glycosyltransferase; hexosyltransferase

Query Match 58.8%; Score 40; DB 2; Length 826;
Best Local Similarity 58.3%; Pred. No. 58;
Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 IRRKFTSEYNEY 13
  ||| : |||
Db 421 INNAFTGDYNEY 432

RESULT 4
B90408
hypothetical protein SSO2370 [imported] - Sulfolobus solfataricus
C:Species: Sulfolobus solfataricus
C>Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 24-May-2001
C:Accession: B90408
R:She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Chan-
Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, P.
arrett, R.A.; Ragan, M.A.; Senses, C.W.; Van der Oost, J.
submitted to GenBank, April 2001
A:Description: Sulfolobus solfataricus complete genome.
A:Reference number: A99139
A:Accession: B90408
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-179 <KUR>
A:Cross-references: GS:AB006641; NID:913815673; PIDN:AAK42521.1; GSPDB:GN00155
C:Genetics:
A:Gene: SSO2370

Query Match 57.4%; Score 39; DB 2; Length 179;
Best Local Similarity 63.6%; Pred. No. 18;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 IRRKFTSEYNEY 12
  ||| : |||
Db 80 VRRKITSVNE 90

RESULT 5
T18268
hypothetical protein - slime mold (Dictyostelium discoideum)
C:Species: Dictyostelium discoideum
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T18268

```

```

R:Ennis, H.L.
submitted to the EMBL Data Library, February 1995
A:Reference number: Z18851
A:Accession: T18268
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-280 <ENN>
A:Cross-references: EMBL:U20608; NID:9676858; PID:92081632; PIDN:AAB54082.1
C:Genetics:
A:Introns: 42/1

Query Match 57.4%; Score 39; DB 2; Length 280;
Best Local Similarity 46.2%; Pred. No. 28;
Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 LIRKFTSEYNEY 13
  ||| : |||
Db 31 LVKTNFTIEYNKF 43

RESULT 6
C89966
hypothetical protein [imported] - Staphylococcus aureus (strain N315)
C:Species: Staphylococcus aureus
C>Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
C:Accession: C89966
R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguch-
ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.; I-
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiratsugu, K.
Lauret 357, 1225-1240, 2001
A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.
A:Reference number: A89758; MUID:21311952; PMID:11418146
A:Accession: C89966
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-190 <KUR>
A:Cross-references: GB:BA000018; PID:gl3701596; PIDN:BA842889.1; GSPDB:GN00149
A:Experimental source: strain N315
C:Genetics:
A:Gene: SA1621

Query Match 55.9%; Score 38; DB 2; Length 190;
Best Local Similarity 45.5%; Pred. No. 29;
Matches 5; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 3 RRRKFTSEYNEY 13
  : : : : :
Db 3 KNEYTAKYNEY 13

RESULT 7
E90176
conserved hypothetical protein [imported] - Sulfolobus solfataricus
C:Species: Sulfolobus solfataricus
C>Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 24-May-2001
C:Accession: E90176
R:She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Chan-
Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, P.
arrett, R.A.; Ragan, M.A.; Senses, C.W.; Van der Oost, J.
submitted to GenBank, April 2001
A:Description: Sulfolobus solfataricus complete genome.
A:Reference number: A99139
A:Accession: E90176
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-337 <KUR>
A:Cross-references: GB:AB006641; NID:913813480; PIDN:AAK40668.1; GSPDB:GN00155
C:Genetics:
A:Gene: SSO0335

Query Match 55.9%; Score 38; DB 2; Length 337;
Best Local Similarity 66.7%; Pred. No. 52;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

```

QY 5 KFTSEYNEY 13  
 Db 49 RYTSAYNEY 57

## RESULT 8

H97107  
 uncharacterized protein from YceG family [imported] - Clostridium acetobutylicum  
 C:Species: Clostridium acetobutylicum 14-Sep-2001 #text\_change 01-Mar-2002  
 C:Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 01-Mar-2002  
 C:Accession: H97107  
 R:Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee, J.; Dally, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.  
 J. Bacteriol. 183, 4823-4838, 2001  
 A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clostridium acetobutylicum  
 A:Reference number: A96900, MUID:21359325; PMID:21359325  
 A:Accession: H97107  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-339 <KUR>  
 A:Cross-references: GB:AE001437; PIDN:AAK79651.1; PID:G15024648; GSPDB:GN00168  
 A:Experimental source: Clostridium acetobutylicum ATCC824  
 C:Genetics:  
 A:Gene: CAC1685  
 C:Superfamily: yceG protein

Query Match 55.9%; Score 38; DB 2; Length 339;  
 Best Local Similarity 75.0%; Pred. No. 52;  
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 6 FTSEYNEY 13  
 Db 322 FTSYNEF 329

## RESULT 9

C72219  
 DNA mismatch repair protein - Thermotoga maritima (strain MSB8)  
 C:Species: Thermotoga maritima  
 C:Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jul-2000  
 C:Accession: C72219  
 R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey, Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.; C.W.  
 Nature 399, 323-329, 1999  
 A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome sequencing of *Thermotoga maritima*  
 A:Reference number: A72200; MUID:99287316; PMID:10360571  
 A:Accession: C72219  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-793 <ARN>  
 A:Cross-references: GB:AE001811; GB:AE000512; MID:94982291; PIDN:AAK36785.1; PID:G498229  
 A:Experimental source: strain MSB8  
 C:Genetics:  
 A:Gene: TM1719  
 C:Superfamily: DNA mismatch repair protein muts

Query Match 55.9%; Score 38; DB 2; Length 793;  
 Best Local Similarity 53.8%; Pred. No. 1.3e+02;  
 Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 LIRKFTSEYNEY 13  
 Db 401 VIKRFSSELDY 413

## RESULT 10

A33284  
 myosin heavy chain IB - slime mold (Dictyostelium discoideum)  
 N:Contains: myosin Apase (SC 3.6.4.1)  
 C:Species: Dictyostelium discoideum  
 C:Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 19-Apr-2002

C:Accession: A33284  
 R:Jung, G.; Saxe III, C.L.; Kimmel, A.R.; Hammer III, J.A.  
 Proc. Natl. Acad. Sci. U.S.A. 86, 6186-6190, 1989  
 A:Title: Dictyostelium discoideum contains a gene encoding a myosin I heavy chain.  
 A:Reference number: A33284; MUID:89345628; PMID:2762320  
 A:Accession: A33284  
 A:Molecule type: DNA  
 A:Residues: 1-1111 <JUN>  
 A:Cross-references: GB:M26037; MID:G167838; PIDN:AAA33229.1; PID:G-67839  
 C:Genetics:  
 A:Introns: 1/3; 39/3  
 C:Superfamily: protozoan myosin heavy chain IB; myosin motor domain homology; SH3 homology  
 C:Keywords: actin binding; ATP; hydrolase; nucleotide binding; P-loop; phosphoprotein; t  
 F12-678/Domain: myosin motor domain homology <MMOI>  
 F102-109/Region: nucleotide-binding motif A (P-loop)  
 F1566-587/Region: actin binding #status predicted  
 F712-1111/Domain: carboxyl-terminal <CTD>  
 F712-901/Region: basic  
 F722-1058/Region: alanine/glutamine/glycine/proline-rich  
 F1060-1109/Domain: SH3 homology <SH3>  
 F108/Binding site: ATP (lys) #status predicted  
 F332/Binding site: phosphate (Ser) (covalent) #status predicted

Query Match 55.9%; Score 38; DB 1; Length 1111;  
 Best Local Similarity 58.3%; Pred. No. 1.8e+02;  
 Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 2 IRRKFTSEYNEY 13  
 Db 736 IDRKFTSDYDF 747

## RESULT 11

S55939  
 telomerase component p80 - Tetrahymena thermophila  
 C:Species: Tetrahymena thermophila  
 C:Date: 27-Oct-1995 #sequence\_revision 03-Nov-1995 #text\_change 21-Jul-2000  
 C:Accession: S55939  
 R:Collins, K.; Kobayashi, R.; Greider, C.W.  
 Cell 81, 677-686, 1995  
 A:Title: Purification of Tetrahymena telomerase and cloning of genes encoding the two pri  
 A:Reference number: S55939; MUID:95292335; PMID:7774009  
 A:Accession: S55939  
 A:Status: preliminary; nucleic acid sequence not shown  
 A:Molecule type: mRNA  
 A:Residues: 1-719 <COI>  
 A:Cross-references: EMBL:U25641; MID:G662483; PIDN:AAC46601.1; PID:G662484  
 C:Genetics:  
 A:Genetic code: SGC5

Query Match 55.1%; Score 37.5; DB 2; Length 719;  
 Best Local Similarity 66.7%; Pred. No. 1.4e+02;  
 Matches 8; Conservative 2; Mismatches 1; Indels 1; Gaps 1;

QY 2 IRRKFTSEYNEY 13  
 Db 177 VRSKF-SEFNEY 187

## RESULT 12

H70484  
 probable CDPdiacylglycerol-glycerol-3-phosphate 3-phosphatidyltransferase (EC 2.7.8.5) -  
 C:Species: Aquifex aeolicus  
 C:Date: 08-May-1998 #sequence\_revision 08-May-1998 #text\_change 03-Jun-2002  
 C:Accession: H70484  
 R:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Ove

Nature 392, 353-358, 1998  
 A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.  
 A:Reference number: A70300; MUID:98196666; PMID:9537320  
 A:Accession: H70484  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA

A;Residues: 1-178 <AQF>  
A;Cross-references: GB:AE000775; NID:g2984342; PIDN:AAO07866.1; PID:g2984346; GB:AE00065  
A;Experimental source: strain VFS  
C;Genetics:  
A;Gene: pgsA1  
C;Superfamily: CDPdiacylglycerol-glycerol-3-phosphate 3-phosphatidytransferase  
C;Keywords: transferase

Query Match 54.4%; Score 37; DB 2; Length 178;  
Best Local Similarity 61.5%; Pred. No. 40;  
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 LIRKFTSEYNE 13  
|||:|||||:  
Db 107 LIRKFTSEYNE 119

RESULT 13  
S56199  
hypothetical protein YFL056c - yeast (*Saccharomyces cerevisiae*)  
N;Alternate names: hypothetical protein R015  
C;Species: *Saccharomyces cerevisiae*  
C;Date: 02-Sep-1995 #sequence\_revision 19-Oct-1995 #text\_change 19-Apr-2002  
C;Accession: S56199; S62281  
R;Murakami, Y.; Naitou, M.; Hagiwara, H.; Shibata, T.; Ozawa, M.; Sasanuma, S.I.; Sasano  
submitted to the EMBL Data Library, May 1995  
A;Description: Analysis of the nucleotide sequence of chromosome VI from *Saccharomyces ce*  
A;Reference number: S56186  
A;Accession: S56199  
A;Molecule type: DNA  
A;Residues: 1-212 <MUR>  
A;Cross-references: EMBL:D50617; NID:g836685; PIDN:BA09185.1; PID:g836699; MIPS:YFL056c  
R;Murakami, Y.  
submitted to the EMBL Data Library, December 1994  
A;Reference number: S62230  
A;Accession: S62281  
A;Molecule type: DNA  
A;Residues: 1-212 <MUR>  
A;Cross-references: EMBL:D44603; NID:g871957; PIDN:BA08063.1; PID:g871970  
C;Genetics:  
A;Gene: SGD:AAD6  
A;Cross-references: SGD:S0001838  
A;Map position: 6L  
C;Superfamily: fission yeast pyridoxine 4-dehydrogenase

Query Match 54.4%; Score 37; DB 2; Length 212;  
Best Local Similarity 46.2%; Pred. No. 48;  
Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 LIRKFTSEYNE 13  
|||:|||||:  
Db 99 VIATKFTDYKY 111

RESULT 14  
A82668  
molybdopterin biosynthesis protein XF1545 [imported] - *Xylella fastidiosa* (strain 95a5c)  
C;Species: *Xylella fastidiosa*  
C;Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 20-Aug-2000  
C;Accession: A82668  
R;Anonymous, The *Xylella fastidiosa* Consortium of the Organization for Nucleotide Sequen  
Nature 406, 151-157, 2000  
A;Title: The genome sequence of the plant pathogen *Xylella fastidiosa*.  
A;Reference number: A82515; MUID:20365717; PMID:10910347  
A;Note: for a complete list of authors see reference number A59328 below  
A;Accession: A82668  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-276 <SIM>  
A;Cross-references: GB:AE003984; GB:AE003849; NID:g9106581; PIDN:AAF84354.1; GSPDB:GN001  
R;Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A  
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H

as-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.  
submitted to GenBank, June 2000  
A;Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohma  
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuranae, E.B.; Laigra  
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E.  
A;Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;  
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A.  
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak;  
A;Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveira  
M.; Tsunako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z  
A;Reference number: A59328  
A;Contents: annotation  
C;Genetics:  
A;Gene: XF1545

Query Match 54.4%; Score 37; DB 2; Length 276;  
Best Local Similarity 63.6%; Pred. No. 63;  
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 LIRKFTSEYNE 11  
|||:|||||:  
Db 188 LIRKFTSEYNE 198

RESULT 15  
B90280  
galactarate dehydratase, probable [gard] [imported] - *Sulfolobus solfataricus*  
C;Species: *Sulfolobus solfataricus*  
C;Date: 24-May-2001 #sequence\_revision 24-May-2001 #text\_change 24-May-2001  
C;Accession: B90280  
R;She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Chan-v  
Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, P.  
arrett, R.A.; Ragan, M.A.; Sersen, C.W.; Van der Oost, J.  
submitted to GenBank, April 2001  
A;Description: *Sulfolobus solfataricus* complete genome.  
A;Reference number: A99139  
A;Accession: B90280  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-387 <KUR>  
A;Cross-references: GB:AE006641; NID:g13814453; PIDN:AAK41497.1; GSPDB:GN00155  
C;Genetics:  
A;Gene: gard

Query Match 54.4%; Score 37; DB 2; Length 387;  
Best Local Similarity 53.8%; Pred. No. 90;  
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 LIRKFTSEYNE 13  
|||:|||||:  
Db 203 VLKDFMRYNE 215

Search completed: October 5, 2004, 08:16:58  
Job time : 6.10648 secs



P40009 saccharomyc  
Q92844 listeria in  
Q8y4f5 listeria mo  
Q971q2 clostridium  
P17794 agrobacteri  
Q9zcn6 rickettsia  
P54112 staphylococ  
P38811 saccharomyc  
P36864 volvox cart  
Q39573 chlamydomon  
P50754 xenopus lae  
P96796 methanosarc

34 35 51.5 630 1 YND1\_YEAST  
35 35 51.5 658 1 UVRB\_LISIN  
36 35 51.5 658 1 UVRB\_LISMO  
37 35 51.5 666 1 UVRB\_CLOAB  
38 35 51.5 789 1 VIB4\_AGR5  
39 35 51.5 814 1 SVV\_RICPR  
40 35 51.5 893 1 GYRA\_STAEP  
41 35 51.5 3744 1 TRAI\_YEAST  
42 34 50.0 205 1 YPT5\_VOLCA  
43 34 50.0 206 1 YPT5\_CHLRE  
44 34 50.0 243 1 TRIC\_XENLA  
45 34 50.0 263 1 HDRE\_METBA

ALIGNMENTS

RESULT 1  
UVRB2 MYCPU  
ID UVRB2 MYCPU STANDARD; PRT; 660 AA.  
AC Q92B21;  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE UVRABC system protein B (UvrB protein) (Excinuclease ABC subunit B).  
GN UVRB  
OS Mycoplasma pulmonis.  
CC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.  
OX NCBI\_taxID=2107;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=X1048;  
RA Minion F.C., Jarvill-Taylor K.J.;  
RT "Molecular analysis of the uvrB homologue in Mycoplasma pulmonis.";  
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: The UvrABC repair system catalyzes the recognition and  
processing of DNA lesions. A damage recognition complex composed  
of 2 uvrA and 2 uvrB subunits scans DNA for abnormalities. Upon  
binding of the uvrA(2)B(2) complex to a putative damaged site, the  
DNA wraps around one uvrB monomer. DNA wrap is dependent on ATP  
binding by uvrB and probably causes local melting of the DNA  
strands. Then uvrB probes one DNA strand for the presence of a  
lesion. If a lesion is found the uvrA subunits dissociate and the  
uvrB-DNA preincision complex is formed. This complex is  
subsequently bound by uvrC and the second uvrB is released. If no  
lesion is found the DNA wraps around the other uvrB subunit that  
will check the other stand for damage (By similarity).  
CC -!- SUBUNIT: Forms a heterotrimer with uvrA during the search for  
lesions. Interacts with uvrC in an incision complex (By  
similarity).  
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
CC -!- DOMAIN: The beta-hairpin motif is involved in DNA binding (By  
similarity).  
CC -!- SIMILARITY: Belongs to the uvrB family.  
CC -!- SIMILARITY: Contains 1 UVR domain.

This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL: U59874; AAD00088.1; --.  
HSSP: P56981; 1D9X.  
HMAP: NF\_00204; --; 1.  
InterPro: IPR001410; DEAD.  
InterPro: IPR001650; Helicase\_C.  
InterPro: IPR004807; UvrB.  
InterPro: IPR001943; UvrB/C.  
Pfam: PF00270; DEAD; 1.

34 35 51.5 630 1 YND1\_YEAST  
35 35 51.5 658 1 UVRB\_LISIN  
36 35 51.5 658 1 UVRB\_LISMO  
37 35 51.5 666 1 UVRB\_CLOAB  
38 35 51.5 789 1 VIB4\_AGR5  
39 35 51.5 814 1 SVV\_RICPR  
40 35 51.5 893 1 GYRA\_STAEP  
41 35 51.5 3744 1 TRAI\_YEAST  
42 34 50.0 205 1 YPT5\_VOLCA  
43 34 50.0 206 1 YPT5\_CHLRE  
44 34 50.0 243 1 TRIC\_XENLA  
45 34 50.0 263 1 HDRE\_METBA

ALIGNMENTS

RESULT 1  
UVRB2 MYCPU  
ID UVRB2 MYCPU STANDARD; PRT; 660 AA.  
AC Q92B21;  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE UVRABC system protein B (UvrB protein) (Excinuclease ABC subunit B).  
GN UVRB  
OS Mycoplasma pulmonis.  
CC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.  
OX NCBI\_taxID=2107;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=X1048;  
RA Minion F.C., Jarvill-Taylor K.J.;  
RT "Molecular analysis of the uvrB homologue in Mycoplasma pulmonis.";  
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: The UvrABC repair system catalyzes the recognition and  
processing of DNA lesions. A damage recognition complex composed  
of 2 uvrA and 2 uvrB subunits scans DNA for abnormalities. Upon  
binding of the uvrA(2)B(2) complex to a putative damaged site, the  
DNA wraps around one uvrB monomer. DNA wrap is dependent on ATP  
binding by uvrB and probably causes local melting of the DNA  
strands. Then uvrB probes one DNA strand for the presence of a  
lesion. If a lesion is found the uvrA subunits dissociate and the  
uvrB-DNA preincision complex is formed. This complex is  
subsequently bound by uvrC and the second uvrB is released. If no  
lesion is found the DNA wraps around the other uvrB subunit that  
will check the other stand for damage (By similarity).  
CC -!- SUBUNIT: Forms a heterotrimer with uvrA during the search for  
lesions. Interacts with uvrC in an incision complex (By  
similarity).  
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
CC -!- DOMAIN: The beta-hairpin motif is involved in DNA binding (By  
similarity).  
CC -!- SIMILARITY: Belongs to the uvrB family.  
CC -!- SIMILARITY: Contains 1 UVR domain.

This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL: U59874; AAD00088.1; --.  
HSSP: P56981; 1D9X.  
HMAP: NF\_00204; --; 1.  
InterPro: IPR001410; DEAD.  
InterPro: IPR001650; Helicase\_C.  
InterPro: IPR004807; UvrB.  
InterPro: IPR001943; UvrB/C.  
Pfam: PF00270; DEAD; 1.

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2004, 07:29:35 ; Search time 1.16358 Seconds  
(without alignments)  
581.749 Million cell updates/sec

Title: US-09-805-290A-12  
Perfect score: 68  
Sequence: 1 LIRKFTSEYNEY 13

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	42	61.8	660	UVRB2 MYCPU	Q92B21 mycoplasma
2	39	57.4	293	SCRK_STRMU	Q07211 streptococ
3	38	55.9	663	UVRB_FUSNN	Q8rgt2 fusobacteri
4	38	55.9	793	MUTS_THEMEA	P74926 thetomotoga
5	38	55.9	1111	MYSE_DICDI	P34092 dictyosteli
6	37.5	55.1	719	TE80_TETTH	Q24818 tetrahymena
7	37	54.4	178	PGSA_AQUAE	O67908 aquifex aeo
8	37	54.4	212	AAD6_YEAST	P43547 saccharomyc
9	37	54.4	500	VLI_HPVA4	P50816 human papil
10	37	54.4	501	VLI_HPVS5	P50820 human papil
11	37	54.4	578	NH35_CABEL	Q17771 caenorhabdi
12	37	54.4	669	UVRB_SYNY3	O55170 synschoyest
13	37	54.4	861	GLGB_SOLTU	P30924 solanum tub
14	36	52.9	329	AAD4_YEAST	Q07747 saccharomyc
15	36	52.9	376	AADE_YEAST	P42884 saccharomyc
16	36	52.9	394	CARA_DEIRA	Q9rw14 deinococcus
17	36	52.9	592	SYD_FUSNN	O8rg14 fusobacteri
18	36	52.9	662	UVRB_THETN	O8rm4 thermoaer
19	36	52.9	874	SYLM_SCHPO	Q09828 schizosacch
20	36	52.9	1534	RHBI_CANPA	Q28298 canis fami
21	35	51.5	184	YGB8_YEAST	P53193 saccharomyc
22	35	51.5	217	RNH2_CHLMU	Q9p110 chlamydia m
23	35	51.5	243	SDGF_RAT	P24338 rattus norv
24	35	51.5	410	TDP1_HUMAN	Q14186 homo sapien
25	35	51.5	410	TDP1_MOUSE	O08639 mus musculu
26	35	51.5	415	REPI_KLULA	P33775 kluyveromyc
27	35	51.5	415	Y207_METUA	Q80289 methanococ
28	35	51.5	446	TDP2_HUMAN	Q14188 homo sapien
29	35	51.5	446	TDP2_MOUSE	Q64163 mus musculu
30	35	51.5	474	CLND_LACLA	Q00565 lactococcus
31	35	51.5	482	CLS2_BACSU	P71040 bacillus su
32	35	51.5	517	MTC1_COSRQ	P42828 corynebacte
33	35	51.5	617	RNC1_YEAST	P33753 saccharomyc

```

DR Pfam; PF00271; helicase_C; 1.
DR SMART; SM00487; DXDC; 1.
DR SMART; SM00490; HELIC; 1.
DR TIGRFAMs; TIGR00631; uvrB; 1.
DR PROSITE; PS0151; UVR; 1.
KW SOS response; Excision nuclease; DNA repair; DNA recombination;
KW DNA excision; ATP-binding.
FT NP BIND 39 46 ATP (POTENTIAL).
FT DOMAIN 92 115 BETA-HAIRPIN.
FT DOMAIN 622 657 UVR.
SQ SEQUENCE 660 AA; 75884 MW; 33A5AF375705367D CRC64;

Query Match 61.8%; Score 42; DB 1; Length 660;
Best Local Similarity 66.7%; Pred. No. 7; 1; Mismatches 3; Indels 0; Gaps 0;
Matches 8; Conservative 1;

QY 1 LIRKFTSEYNE 12
DB 569 LFKRKIQSEYNE 580

RESULT 2
SCRK_STRMU STANDARD; PRT; 293 AA.
AC Q07211;
DT 01-OCT-1994 (Rel. 30, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE 28-FEB-2003 (Rel. 41, Last annotation update)
DE Fructokinase [EC 2.7.1.4].
GN SCRK OR SMU.1840.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GS-5;
RA Sato Y., Yamamoto Y., Kizaki H., Kuramitsu H.K.;
RX MEDLINE=22295063; PubMed=8336109;
RT "Isolation, characterization and sequence analysis of the scrk gene
RT encoding fructokinase of Streptococcus mutans.";
RL J. Gen. Microbiol. 139:921-927(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=UAI59 / ATCC 700610 / Serotype C;
RX MEDLINE=22295063; PubMed=12397186;
RA Ajdic D., McShan W.M., McLaughlin R.E., Savic G., Chang J.,
RA Carson M.B., Primeaux C., Tian R., Kenton S., Jia H., Lin S., Qian Y.,
RA Li S., Zhu H., Najjar F., Lai H., White J., Roe B.A., Ferretti J.J.;
RT "Genome sequence of Streptococcus mutans UAI59, a cariogenic dental
RT pathogen.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439(2002).
CC -!- CATALYTIC ACTIVITY: ATP + D-fructose = ADP + D-fructose 6-
CC phosphate.
CC -!- SIMILARITY: BELONGS TO THE ROK (NAGC/XYLR) FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D13175; BAA02467.1; -.
CC EMBL; AE015011; AAN59463.1; -.
CC InterPro; IPR000600; ROK.
CC Pfam; PF00480; ROK; 1.
CC PROSITE; PS01125; ROK; FALSE NEG.
KW Transferase; Kinase; Complete proteome.
FT CONFLICT 139 139 T -> A (IN REF. 1).
FT CONFLICT 139 139
SQ SEQUENCE 293 AA; 31712 MW; 1089AF37B2C5B807 CRC64;

Query Match 57.4%; Score 39; DB 1; Length 293;
Best Local Similarity 58.3%; Pred. No. 10;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 IRRKFTSEYNE 13
DB 247 VREKFTSLNDY 258

RESULT 3
UVRB_FUSNN STANDARD; PRT; 663 AA.
AC Q8GR2;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE UVRABC system protein B (UvrB protein) (Excinuclease ABC subunit B).
GN UVRB OR FN0224.
OS Fusobacterium nucleatum (subsp. nucleatum).
OC Bacteria; Fusobacteriia; Fusobacteriales; Fusobacteriaceae;
OC Fusobacterium.
OX NCBI_TaxID=76856;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 25586;
RX MEDLINE=21886394; PubMed=11889109;
RA Kapural V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A.,
RA Bhattacharyya A., Bartman A., Gardner W., Grechkin G., Zhu L.,
RA Vasileva O., Chu L., Kogan Y., Chaga O., Goltzman E., Bernal A.,
RA Larsen N., D'Souza M., Walunas T., Pusch G., Haselkorn R.,
RA Fongstein M., Kyripides N., Overbeek R.;
RT "Genome sequence and analysis of the oral bacterium Fusobacterium
RT nucleatum strain ATCC 25586.";
RL J. Bacteriol. 184:2005-2018(2002).
CC -!- FUNCTION: The UVRABC repair system catalyzes the recognition and
CC processing of DNA lesions. A damage recognition complex composed
CC of 2 uvrA and 2 uvrB subunits scans DNA for abnormalities. Upon
CC binding of the uvrA(2)B(2) complex to a putative damaged site, the
CC DNA wraps around one uvrB monomer. DNA wrap is dependent on ATP
CC binding by uvrB and probably causes local melting of the DNA
CC helix, facilitating insertion of uvrB beta-hairpin between the DNA
CC strands. Then uvrB probes one DNA strand for the presence of a
CC lesion. If a lesion is found the uvrA subunits dissociate and the
CC uvrB-DNA preincision complex is formed. This complex is
CC subsequently bound by uvrC and the second uvrB is released. If no
CC lesion is found, the DNA wraps around the other uvrB subunit that
CC will check the other strand for damage (By similarity).
CC -!- SUBUNIT: Forms a heterotrimer with uvrA during the search for
CC lesions. Interacts with uvrC in an incision complex (By
CC similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- DOMAIN: The beta-hairpin motif is involved in DNA binding (By
CC similarity).
CC -!- SIMILARITY: Belongs to the uvrB family.
CC -!- SIMILARITY: Contains 1 UVR domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE010535; AAL94430.1; -.
CC HAMAP; MF_00204; -.
CC InterPro; IPR001410; DEAD.
CC InterPro; IPR001650; Helicase_C.
CC InterPro; IPR004807; UvrB.
CC InterPro; IPR001943; UvrB/C.
CC Pfam; PF00270; DEAD; 1.
CC Pfam; PF00271; helicase_C; 1.

```

```
DR Pfam: PF02151; UVR: 1.
DR SMART: SM00487; DEXDC: 1.
DR SMART: SM00490; HELIC: 1.
DR TIGRFAMs: TIGR00631; uvrB: 1.
DR PROSITE: PS0151; UVR: 1.
DR SOS response; Excision nuclease; DNA repair; DNA recombination;
KW DNA excision; ATP-binding; Complete proteome.
FT NP BIND 40 47
FT DOMAIN 93 116
FT DOMAIN 624 659
FT SEQUENCE 663 AA; 76995 MW; DEF00752C19D1389 CRC64;

Query Match 55.9%; Score 38; DB 1; Length 663;
Best Local Similarity 63.6%; Pred. No. 38;
Matches 7; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 RRKPTSEYNEY 13
DB 572 RRKIQEYNAY 582

RESULT 4
MUTS_THEME STANDARD; PRT; 793 AA.
AC P74926;
DT 15-JUL-1998 (Rel. 36, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE DNA mismatch repair protein mutS.
MUTS OR TM1719.
OS Thermotoga maritima.
OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
OX NCBI_TaxID=2336;
RN [1]
RP SEQUENCE FROM N.A.
RA Metmur J.G., Rosenfeld A., Wong D.M.;
RT "Hyperthermophilic MutS proteins: isolation, characterization and
enhancement of PCR specificity.";
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=MSB8 / DSM 3109 / ATCC 43589;
RX MEDLINE=99287315; PubMed=10360571;
RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RT "Evidence for lateral gene transfer between Archaea and Bacteria from
genome sequence of Thermotoga maritima.";
RL Nature 399:323-329(1999).
CC -!- FUNCTION: This protein is involved in the repair of mismatches in
DNA. It is possible that it carries out the mismatch recognition
step. This protein has a weak ATPase activity (By similarity).
CC -!- SIMILARITY: Belongs to the DNA mismatch repair mutS family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
CC EMBL; U71155; AAB1999.1; -.
CC EMBL; AE001811; AAD36785.1; -.
CC PIR; C72219; C72219.
CC TIGR; TM1719; -.
CC HAVAP; MF 00096; -.
DR InterPro; IPR005748; MutS1.
DR InterPro; IPR000432; MutS_C.
DR InterPro; IPR007860; MutS-II.
```

```
DR InterPro; IPR007696; MutS-III.
DR InterPro; IPR007861; MutS-IV.
DR InterPro; IPR007695; MutS_N.
DR Pfam; PF01624; MutS_I; 1.
DR Pfam; PF05188; MutS-II; 1.
DR Pfam; PF05192; MutS-III; 1.
DR Pfam; PF05190; MutS-IV; 1.
DR Pfam; PF00488; MutS_V; 1.
DR ProDom; PD001263; MutS_C; 1.
DR SMART; SM00534; MutSact; 1.
DR SMART; SM00533; MutSd; 1.
DR TIGRFAMs; TIGR01070; mutS1; 1.
DR PROSITE; PS00486; DNA_MISMATCH_REPAIR_2; 1.
KW DNA repair; ATP-binding; DNA-binding; Complete proteome.
FT NP BIND 589 596
FT CONFLICT 233 233 A -> G (IN REF. 1).
FT CONFLICT 262 262 L -> W (IN REF. 1).
FT CONFLICT 287 287 L -> G (IN REF. 1).
FT CONFLICT 506 506 K -> T (IN REF. 1).
FT SEQUENCE 793 AA; 91065 MW; 1ECB2342E4F9B1ED CRC64;

Query Match 55.9%; Score 38; DB 1; Length 793;
Best Local Similarity 53.8%; Pred. No. 46;
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 LIRKPTSEYNEY 13
DB 401 VIKRGFSSELDEY 413

RESULT 5
MUTS_DICDI STANDARD; PRT; 1111 AA.
AC P34092;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Myosin IB heavy chain.
GN MYOB OR MYOA OR MYIB.
OS Dictyostelium discoideum (Slime mold). Dictyostelium.
OC Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AX3;
RX MEDLINE=89345628; PubMed=2762320;
RA Jung G., Saxe C.L. III, Kimmel A.R., Hammer J.A. III;
RT "Dictyostelium discoideum contains a gene encoding a myosin I heavy
chain."
RL Proc. Natl. Acad. Sci. U.S.A. 86:6186-6190(1989).
RN [2]
RP SEQUENCE OF 481-490; 656-666 AND 783-798.
RC STRAIN=AX3;
RX MEDLINE=93315475; PubMed=8325874;
RA Jung G., Fukui Y., Martin B., Hammer J.A. III;
RT "Sequence, expression pattern, intracellular localization, and
targeted disruption of the Dictyostelium myosin ID heavy chain
isoform."
RL J. Biol. Chem. 268:14981-14990(1993).
CC -!- FUNCTION: Myosin is a protein that binds to actin and has ATPase
activity that is activated by actin. Myosin IB may have a role in
chemotaxis and aggregation; it could serve to stabilize and even
retract cortical structures, such as pseudopods and lamellopods.
CC -!- SUBUNIT: Myosin I heavy chain is single-headed. Dimer of a heavy
and a light chain. Inability to self-assemble into filaments.
CC -!- SUBCELLULAR LOCATION: Highest concentration just beneath the
plasma membrane in the anterior pseudopod at the leading edge of
the cell.
CC -!- DOMAIN: TH.1 binds directly to anionic phospholipid membranes;
myosins I could therefore move actin relative to membranes and
vice versa. TH.2 and SH3 bind tightly to F-actin; this together
with the nucleotide-sensitive site in the head, allows single
molecules of myosin I to cross-link actin filaments.
```

```

CC CC -!- SIMILARITY: Contains 1 myosin-like globular head domain.
CC CC -!- SIMILARITY: Contains 1 SH3 domain.
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
CC CC EMBL; M26037; AAA3229.1; -.
CC CC PIR; A33284; A33284.
CC CC RSP; P08799; 1MND.
CC CC DictyBase; DDB0002013; myoB.
CC CC InterPro; IPR000048; IQ region.
CC CC InterPro; IPR001609; myosin_head.
CC CC InterPro; IPR001452; SH3.
CC CC Pfam; PF00612; IQ; 1.
CC CC Pfam; PF00663; myosin head; 1.
CC CC Pfam; PF00018; SH3; 1.
CC CC PRINTS; PR00193; MYOSINHEAVY.
CC CC PRINTS; PR00452; SH3DOMAIN.
CC CC ProDom; PD000355; myosin head; 1.
CC CC ProDom; PD000666; SH3; 1.
CC CC SMART; SM00242; MYSC; 1.
CC CC SMART; SM00326; SH3; 1.
CC CC PROSITE; PS00002; SH3; 1.
CC CC Chemotaxis; Phosphorylation.
CC CC Myosin; Actin-binding; ATP-binding; SH3 domain; Multigene family;
CC CC -----
CC CC FT DOMAIN 1 634 MYOSIN HEAD-LIKE
CC CC FT DOMAIN 695 921 TAIL HOMOLOG REGION 1 (TH.1).
CC CC FT DOMAIN 922 1052 GLY/PRO/ALA-RICH (TH.2).
CC CC FT DOMAIN 1053 1111 SH3.
CC CC FT NP_BIND 102 109 ATP (POTENTIAL).
CC CC FT DOMAIN 547 627 ACTIN-BINDING.
CC CC FT DOMAIN 951 1015 ASN-RICH.
CC CC FT MOD_RES 332 332 PHOSPHORYLATION (BY SIMILARITY).
CC CC FT SEQUENCE 1111 AA; 124313 MW; CD6386F08DC5642F CRC64;
CC CC -----
CC CC Query Match 55.9%; Score 38; DB 1; Length 1111;
CC CC Best Local Similarity 58.3%; Pred. No. 65;
CC CC Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
CC CC -----
CC CC QY 2 IRRKFTSEYNEY 13
CC CC Db 736 IDRKFTSDYDF 747
CC CC -----
CC CC RESULT 6
CC CC TE80 TETH STANDARD; PRT; 719 AA.
CC CC AC Q94818.
CC CC DT 01-NOV-1997 (Rel. 35; Created)
CC CC DT 01-NOV-1997 (Rel. 35; Last sequence update)
CC CC DT 15-JUL-1998 (Rel. 36; Last annotation update)
CC CC DE Telomerase component p80 (BC 2.7.7.-).
CC CC OS Tetrahymena thermophila.
CC CC OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Hymenostomatida;
CC CC OC Tetrahymenina; Tetrahymena.
CC CC OX NCBI_TaxID=5911;
CC CC [1]
CC CC SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
CC CC MEDLINE=95292335; PubMed=7774009;
CC CC RA Collins K., Kobayashi R., Greider C.W.;
CC CC RT "Purification of Tetrahymena telomerase and cloning of genes encoding
CC CC the two protein components of the enzyme."
CC CC RL Cell 81:677-686(1995).
CC CC -!- FUNCTION: Ribonucleoprotein DNA polymerase that catalyzes the de
CC CC novo synthesis of telomeric simple sequence repeats. P80 binds
CC CC tightly and specifically to the telomerase RNA suggesting its
CC CC association with a region of RNA secondary structure.
CC CC -!- SUBUNIT: Telomerase consist of two subunit, p80 and p95 that form

```

```

CC CC a 1:1:1 complex with the 159 nt telomerase RNA.
CC CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
CC CC EMBL; U25641; AAC46601.1; -.
CC CC PIR; S55939; S55939.
CC CC InterPro; IPR008858; TROVE.
CC CC Pfam; PF05731; TROVE; 1.
CC CC KX Transferase; RNA-directed DNA polymerase; Telomere; Nuclear protein;
CC CC KW DNA-binding.
CC CC SQ SEQUENCE 719 AA; 82351 MW; 8A945A71189CA99C CRC64;
CC CC -----
CC CC Query Match 55.1%; Score 37.5; DB 1; Length 719;
CC CC Best Local Similarity 66.7%; Pred. No. 51;
CC CC Matches 8; Conservative 2; Mismatches 1; Indels 1; Gaps 1;
CC CC -----
CC CC QY 2 IRRKFTSEYNEY 13
CC CC Db 177 VRSKP-SEFNEY 187
CC CC -----
CC CC RESULT 7
CC CC PGSA AQUAE STANDARD; PRT; 178 AA.
CC CC ID PGSA AQUAE
CC CC AC O67908;
CC CC DT 30-MAY-2000 (Rel. 39; Created)
CC CC DT 30-MAY-2000 (Rel. 39; Last sequence update)
CC CC DT 16-OCT-2001 (Rel. 40; Last annotation update)
CC CC DE CDP-diacylglycerol-glycerol-3-phosphate 3-phosphatidyltransferase
CC CC DE (BC 2.7.8.5) (Phosphatidylglycerophosphate synthase) (PGP synthase).
CC CC GN PGSA OR AQ_2154.
CC CC OS Aquifex aeolicus.
CC CC OC Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.
CC CC OX NCBI_TaxID=63363;
CC CC RN [1]
CC CC RP SEQUENCE FROM N.A.
CC CC RC STRAIN=VF5;
CC CC RX MEDLINE=98196866; PubMed=9537320;
CC CC RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
CC CC Graham D.E., Overbeek R., Sneed M.A., Keller M., Aujay M., Huber R.,
CC CC Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
CC CC RT "The complete genome of the hyperthermophilic bacterium Aquifex
CC CC aeolicus."
CC CC RL Nature 392:353-358(1998).
CC CC -!- FUNCTION: This protein catalyzes the committed step to the
CC CC synthesis of the acidic phospholipids (By similarity).
CC CC -!- CATALYTIC ACTIVITY: CDP-diacylglycerol + glycerol 3-phosphate =
CC CC CMP + 3-(3-phosphatidyl)-glycerol 1-phosphate.
CC CC -!- PATHWAY: Acidic phospholipids biosynthesis.
CC CC -!- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).
CC CC -!- SIMILARITY: Belongs to the CDP-alcohol phosphatidyltransferase
CC CC class-I family.
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
CC CC EMBL; AB000775; AAC07865.1; -.
CC CC PIR; H70484; H70484.
CC CC InterPro; IPR000462; CDP-OH P trans.
CC CC Pfam; PF01066; CDP-OH P trans; 1.
CC CC PROSITE; PS00379; CDP_ALCOHOL_P_TRANSF; 1.

```

```
; APPLICANT: IWAMOTO, MASAO
; TITLE OF INVENTION: PROMOTOR SEQUENCE WHICH IS EXPRESSED IN ANTHERS AND IN
; TITLE OF INVENTION: POLLEN
; FILE REFERENCE: YAM 2 0006 US / AR010
; CURRENT APPLICATION NUMBER: US/09/700,971
; CURRENT FILING DATE: 2000-12-08
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 2
; LENGTH: 4670
; TYPE: DNA
; ORGANISM: Oryza sativa
US-09-700-971-2

Query Match      6.6%; Score 124.6; DB 4; Length 4670;
Best Local Similarity 82.8%; Pred. No. 5.4e-26;
Matches 168; Conservative 0; Mismatches 29; Indels 6; Gaps 2;

QY 341 ATGTTTAGACACATGATGGAGTCTAATAATTAATRAAAAAAATAATACACAGATCGT 400
DB 210 ATGTTTGGACATGATGATGATCTAAATGT--AGAAAAAACAATAAACATTTGCG 267
QY 401 GTGTAATTCGAGATAAATCTTTTAAGCCTAATGTCCTCATGA---ACAATGCGTGT 456
DB 268 CTTGAAATACGAGACAAATCTTTTAAGCCTAATGTCCTCATGA---ACAATGCGTGT 327
QY 457 TACAGTAATTCGAGATAAATCTTTTAAGCCTAATGTCCTCATGA---ACAATGCGTGT 516
DB 328 TACATTAATTAATGCTAATGACATTAATTAAGCCTAATGTCCTCATGA---ACAATGCGTGT 387
QY 517 CAGGTGAATATGTAATTTATTT 539
DB 388 CAGACGGAATCTGTAATTTATTT 410

RESULT 3
US-09-807-897-4
; Sequence 4, Application US/09807897
; Patent No. 6660851
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Yoshimitsu
; APPLICANT: Ueki, Jun
; TITLE OF INVENTION: NOVEL DNA FRAGMENT ENHANCING GENE EXPRESSION LEVEL
; FILE REFERENCE: 0230-0156P
; CURRENT APPLICATION NUMBER: US/09/807,897
; CURRENT FILING DATE: 2001-04-19
; PRIOR APPLICATION NUMBER: PCT/JP00/05539
; PRIOR FILING DATE: 2000-08-18
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 4
; LENGTH: 5396
; TYPE: DNA
; ORGANISM: Cryza sativa L.
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Variety: IR24
; OTHER INFORMATION: Library name: lambdADASHII genomic library derived from green leaf
; OTHER INFORMATION: genomic DNA
; OTHER INFORMATION: Clone name: RPG106 SacI-Sali 5.4 kb
; NAME/KEY: misc feature
; OTHER INFORMATION: Strandedness: double - Topology: linear
; OTHER INFORMATION: Molecular type: genomic DNA
; OTHER INFORMATION: Feature: nt1-nt5369, nt3335-nt5108; sequences whose promoter activity
; OTHER INFORMATION: was confirmed by GUS
; NAME/KEY: misc feature
; OTHER INFORMATION: nt4964-nt4969: TATA box like sequence
; OTHER INFORMATION: nt4995, 4996, 4997: transcription initiation points of RPC213 gene
; OTHER INFORMATION: determined by primer extension
; OTHER INFORMATION: nt5016-nt5018: the first ATG of RPC213 gene
; NAME/KEY: misc feature
; OTHER INFORMATION: nt5128-nt5130: the second ATG of RPC213 gene
; OTHER INFORMATION: nt5370-nt5372: the third ATG of RPC213 gene
```

```
; OTHER INFORMATION: nt5162-nt5242: intron
; NAME/KEY: misc feature
; OTHER INFORMATION: nt729-nt734, nt2811-nt2816, nt5103-nt5108: BglII restriction site
; OTHER INFORMATION: nt3335-nt3340: HindIII site
US-09-807-897-4

Query Match      5.9%; Score 110.2; DB 4; Length 5396;
Best Local Similarity 68.6%; Pred. No. 1e-21;
Matches 212; Conservative 0; Mismatches 88; Indels 9; Gaps 4;

QY 337 TTTAATGTTTAGACACATGATGGAGTCTAATAATAT---TAATAAAAAAATAATACAC 393
DB 4649 TTGAACATTTGAACACCTGTATRAAGTATTAAATATATATCTAAAAAAAATAATATATAC 4708
QY 394 AGATCGCTGTGTAATTCGAGATAAATCTTTTAAGCCTAATGTCCTCATG---AACAAAT 449
DB 4709 ATATTACGACTAATTTGCAAGACGAATCTTTTAAGCATAATTTGCTCCATGATTAACAAT 4768
QY 450 GTGGTGTGTACAGTAACATTTGCTTAATGACAGATTAATTAGGCTTAATAAATTCGTCCTCA 509
DB 4769 ATAGTGCTACAGTAACATGCTGCTAATGACGATTAATTAGGCTTAATAAATTCGTCCTCA 4828
QY 510 CAGTTTACAGGTGAAATATGTAATTTATTTAATTAAGTCTATATATAATAATCTTTAAAT 569
DB 4829 C-GTTTACTGACGGATTTCTATAATTTGATTTTATTATTAATGCCCAACACCCCATACAAC 4887
QY 570 AGTGACCGGTATATCCGATGGGAGACACGTAATAACTTTTAAACCAAGTCTTAAACACAA 629
DB 4888 ACTCTATATAATA-CTCAATGTGACGTCGCAAAACCTTTAGACACCTGGATGTAACACCA 4946
QY 630 CCTTGTCTTC 638
DB 4947 CTCGTCTCC 4955
```

```
RESULT 4
US-08-475-891A-1/c
; Sequence 1, Application US/08475891A
; Patent No. 5859339
; GENERAL INFORMATION:
; APPLICANT: Ronald, Pamela C.
; APPLICANT: Wang, Guo-Liang
; APPLICANT: Song, Wen-tuang
; TITLE OF INVENTION: Procedures and Materials for Confering
; NUMBER OF INVENTION: Disease Resistance in Plants
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/475,891A
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/373,375
; FILING DATE: 17-JAN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Bastian, Kevin L.
; REGISTRATION NUMBER: 34,774
; REFERENCE/POCKET NUMBER: 02370-058910US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
```



KM Transferrase; Phospholipid biosynthesis; Transmembrane;  
 KM Complete proteome.  
 FT TRANSMEM 28 48 POTENTIAL.  
 FT TRANSMEM 88 108 POTENTIAL.  
 FT TRANSMEM 125 145 POTENTIAL.  
 FT TRANSMEM 147 167 POTENTIAL.  
 SQ SEQUENCE 178 AA; 20481 MW; 0ACD28A27956FCF9 CRC64;

Query Match 54.4%; Score 37; DB 1; Length 178;  
 Best Local Similarity 61.5%; Pred. No. 14;  
 Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 LIRKFTSEYNEY 13  
 DB 107 LIRKFTPEPSIX 119

RESULT 8  
 ADE\_YEAST STANDARD; PRT; 212 AA.  
 ID AAD6\_YEAST  
 AC P43547;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Putative aryl-alcohol dehydrogenase AAD6 (EC 1.1.1.1-).  
 GN AAD6 OR YFL056C.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 RX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=8288C / AB972;  
 RX MEDLINE=95400292; PubMed=7670463;  
 RA Murakami Y., Naitou M., Hagiwara H., Shibata T., Ozawa M.,  
 RA Sasanuma S.-I., Sasanuma M., Tsuchiya Y., Soeda E., Yokoyama K.,  
 RA Yamazaki M., Tashiro H., Eki T.;  
 RT "Analysis of the nucleotide sequence of chromosome VI from  
 RT Saccharomyces cerevisiae";  
 RL Nac. Genet. 10:261-268(1995).  
 CC -1- SIMILARITY: BELONGS TO THE ALDO/KETO REDUCTASE 2 FAMILY. STRONG,  
 CC TO THE N-TERMINUS OF YEAST AAD14.

-----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

EMBL; D50617; BAA09185.1; -  
 DR PIR; S56199; S56199.  
 DR Germonline; 140100; -  
 DR SGD; S0001838; AAD6.  
 DR InterPro; IPR001395; Aldo/Ket\_red.  
 DR Pfam; PF00248; aldo\_ket\_red; 1.  
 DR ProDom; PD000288; Aldo/Ket\_red; 1.  
 KW Oxidoreductase.  
 FT ACT SITE 151 151 HYDROGEN-BOND DONOR (PROBABLE).  
 SQ SEQUENCE 212 AA; 23918 MW; C3FF042B12C91580 CRC64;

Query Match 54.4%; Score 37; DB 1; Length 212;  
 Best Local Similarity 46.2%; Pred. No. 17;  
 Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 LIRKFTSEYNEY 13  
 DB 99 VIALKFTDYKRY 111

RESULT 9  
 VIL\_HPV44

ID VIL\_HPV44 STANDARD; PRT; 500 AA.  
 AC P50816; Q80919;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE Major capsid protein L1.  
 GN L1  
 OS Human papillomavirus type 44.  
 OC Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;  
 OC Papillomavirus.  
 OX NCBI\_TaxID=10592;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Delius H.;  
 RL Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.

-----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; U31788; AAA79463.1; -  
 DR EMBL; U12493; AAA67237.1; -  
 DR InterPro; IPR002210; PV\_capsid\_L1  
 DR InterPro; IPR008975; Viral\_cap\_coat.  
 DR Pfam; PF00500; late\_protein\_L1; 1.  
 DR PRINTS; PR00865; HPVCAPSID.L1.  
 DR ProDom; PD000544; PV\_capsid\_L1; 1.  
 KW Coat protein; Late protein.  
 SQ SEQUENCE 500 AA; 55849 MW; 277A4AB7CA421F9 CRC64;

-----  
 CC Query Match 54.4%; Score 37; DB 1; Length 500;  
 CC Best Local Similarity 50.0%; Pred. No. 42;  
 CC Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 2 IIRKFTSEYNEY 13  
 DB 448 LIRKFTSEYDQY 459

RESULT 10  
 VIL\_HPV55 STANDARD; PRT; 501 AA.  
 ID VIL\_HPV55  
 AC P50820; Q80940;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE Major capsid protein L1.  
 GN L1  
 OS Human papillomavirus type 55.  
 OC Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;  
 OC Papillomavirus.  
 OX NCBI\_TaxID=37114;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Delius H.;  
 RL Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE OF 312-462 FROM N.A.  
 RX MEDLINE=95052821; PubMed=7963696;

RA Bernard H.U., Chan S.Y., Manos M.M., Ong C.K., Villa L.L.,  
 RA Delius H., Peyron C.L., Bauer H.W., Wheeler C.M.;  
 RT "Identification and assessment of known and novel human  
 RT papillomaviruses by polymerase chain reaction amplification,  
 RT restriction fragment length polymorphisms, nucleotide sequence, and  
 RT phylogenetic algorithms.";  
 RL J. Infect. Dis. 170:1077-1085 (1994).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: U1791; AAA79484.1; -  
 DR EMBL: U12494; AAA67238.1; -  
 DR InterPro: IPR002210; PV\_capsid\_I1.  
 DR InterPro: IPR008975; Viral\_cap\_coat.  
 DR Pfam: PF00500; late\_protein\_L1.  
 DR PRINTS: PR00665; HPVcapsid1.  
 DR ProDom: PD000544; PV\_capsid\_L1.  
 DR Coat protein; late protein.  
 KW  
 SQ SEQUENCE 501 AA; 55787 MW; 0DBC4D47E552E968 CRC64;  
 Query Match 54.4%; Score 37; DB 1; Length 501;  
 Best Local Similarity 50.0%; Pred. No. 42;  
 Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;  
 QY 2 IRRKTSSEYNEY 13  
 DB 448 IRRKTSSEYNEY 459  
 RESULT 11  
 NH35\_CABEL STANDARD; PRT; 578 AA.  
 ID NH35\_CABEL  
 AC Q17771; Q9GTH7;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Nuclear hormone receptor family member nhr-35.  
 GN NHR-35 OR C07A12.3.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Pelodertinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC Geisel C.;  
 RA STRAIN-Bristol NZ;  
 RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP REVISIONS.  
 RA Waterston R.;  
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE OF 35-578 FROM N.A.  
 RA Bogan A., Manna C.V., Yamamoto K., Cohen F., Sluder A.E.;  
 RT "Caenorhabditis elegans nuclear receptor sequences exhibit biophysical  
 RT compatibility with the ligand-binding domain fold.";  
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.  
 CC -----  
 CC -1- FUNCTION: Orphan nuclear receptor.  
 CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).  
 CC -1- SIMILARITY: Belongs to the nuclear hormone receptor family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: U41542; AAK39150.2; -  
 DR EMBL: AF273781; AAG15130.1; -  
 DR HSSP: P19793; 2NUL  
 DR WormRep: C07A12.3; CE29071.  
 DR InterPro: IPR000536; Hormone\_rec\_lig.  
 DR InterPro: IPR001723; Steroidn\_receptor.  
 DR InterPro: IPR008946; Str\_ncl\_receptor.  
 DR Pfam: PF00104; hormone\_rec; 1.  
 DR Pfam: PF00105; zf-C4; 1.  
 DR PRINTS: PR00398; STROPHOMONER.  
 DR PRINTS: PR00047; STRODPFINGER.  
 DR ProDom: PD000035; Znf\_C4steroid; 1.  
 DR SMART: SM00430; HOL1; 1.  
 DR SMART: SM00399; ZNF\_C4; 1.  
 DR PROSITE: PS00031; NUCLEAR RECEPTOR; 1.  
 KW Receptor; Transcription regulation; DNA-binding; Nuclear protein;  
 KW Zinc-finger.  
 FT DNA\_BIND 77 142 NUCLEAR RECEPTOR-TYPE.  
 FT ZN\_FING 77 97 C4-TYPE.  
 FT ZN\_FING 113 137 C4-TYPE.  
 FT FT 267 302 MISSING (IN REF. 3).  
 KW  
 SQ SEQUENCE 578 AA; 65256 MW; 9868FD4FCE868 CRC64;  
 Query Match 54.4%; Score 37; DB 1; Length 578;  
 Best Local Similarity 75.0%; Pred. No. 43;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 QY 6 FTSEYNEY 13  
 DB 566 FTSEYNEY 573  
 RESULT 12  
 UVRB\_SYNY3 STANDARD; PRT; 669 AA.  
 ID UVRB\_SYNY3  
 AC C55170;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE UVRB system protein B (UVRB protein) (Excinuclease ABC subunit B).  
 GN UVRB OR SL0459.  
 OS Synecocystis sp. (strain PCC 6803).  
 OC Bacteria; Cyanobacteria; Chroococcales; Synecocystis.  
 OX NCBI\_TaxID=1148;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96127529; PubMed=8590279;  
 RA Kaneko T., Tanaka A., Sato S., Kotani H., Sazuka T., Miyajima N.,  
 RA Sugita M., Tabata S.;  
 RT "Sequence analysis of the genome of the unicellular cyanobacterium  
 RT Synecocystis sp. strain PCC6803. I. Sequence features in the 1 Mb  
 RT region from map positions 64% to 92% of the genome.";  
 RL DNA Res. 2:153-166 (1995).  
 CC -----  
 CC -1- FUNCTION: The UVRB repair system catalyzes the recognition and  
 CC processing of DNA lesions. A damage recognition complex composed  
 CC of 2 uvrA and 2 uvrB subunits scans DNA for abnormalities. Upon  
 CC binding of the uvrA(2)B(2) complex to a putative damaged site, the  
 CC DNA wraps around one uvrB monomer. DNA wrap is dependent on ATP  
 CC binding by uvrB and probably causes local melting of the DNA  
 CC helix, facilitating insertion of uvrB beta-hairpin between the DNA  
 CC strands. Then uvrB probes one DNA strand for the presence of a  
 CC lesion. If a lesion is found the uvrA subunits dissociate and the  
 CC uvrB-DNA preincision complex is formed. This complex is  
 CC subsequently bound by uvrC and the second uvrB is released. If no  
 CC lesion is found, the DNA wraps around the other uvrB subunit that  
 CC will check the other strand for damage (By similarity).  
 CC -1- SUBUNIT: Forms a heterotrimer with uvrA during the search for  
 CC lesions. Interacts with uvrC in an incision complex (By  
 CC similarity).  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).



```

CC      -1- DOMAIN: The beta-hairpin motif is involved in DNA binding (By
CC      similarity).
CC      -1- SIMILARITY: Belongs to the uvrb family.
CC      -1- SIMILARITY: Contains 1 UVR domain.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL; D64001; BAAL0309.1; -.
CC      PIR; S74391; S74391.
CC      HSSP; Q56243; 1C4O.
CC      HAMAP; MF_00204; -.
CC      InterPro; IPR001410; DEAD.
CC      InterPro; IPR001650; Helicase_C.
CC      InterPro; IPR004807; UVRB.
CC      InterPro; IPR001943; UVRB/C.
CC      Pfam; PF00251; helicase_C; 1.
CC      Pfam; PF02151; UVR; 1.
CC      SMART; SM00487; DEXDC; 1.
CC      SMART; SM00490; HELIC_C; 1.
CC      TIGRFAMs; TIGR00631; uvrB; 1.
CC      PROSITE; PS00151; UVR; 1.
CC      SOS response; Excision nuclease; DNA repair; DNA recombination;
CC      DNA excision; ATP-binding; Complete proteome.
CC      NP_BIND; 40; 47; ATP (POTENTIAL).
CC      DOMAIN; 93; 116; UVR.
CC      DOMAIN; 628; 663; UVR.
CC      SEQUENCE; 669 AA; 76848 MW; 9C5B245A8A90E3 CRC64;
CC
QY      Query Match          54.4%; Score 37; DB 1; Length 669;
QY      Best Local Similarity 70.0%; Pred. No. 58;
QY      Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY
Db      3 RRRKTSSEYNE 12
QY      ||| |||
QY      571 RRRKTOEYNE 580
QY
RESULT 13
GLGB_SOLUTU STANDARD; PRT; 861 AA.
ID   GLGB_SOLUTU
AC   P30924;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   1,4-alpha-glucan branching enzyme (EC 2.4.1.18) (Starch branching
DE   enzyme) (O-enzyme).
GN   SBE1 OR SBE.
OS   Solanum tuberosum (Potato).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
OC   lamids; Solanales; Solanaceae; Solanum.
OX   NCBI_TaxID=4113;
OX   (1)
RN   SEQUENCE FROM N.A.
RC   STRAIN=cv. Diabella;
RX   MEDLINE=94105324; PubMed=8278528;
RA   Poulsen P., Kreiberg J.D.;
RA   "Starch branching enzyme cDNA from Solanum tuberosum.";
RA   Plant Physiol. 102:1053-1054 (1993).
RN   [2]
RP   SEQUENCE OF 279-527 FROM N.A.
RC   STRAIN=cv. Desiree; TISSUE=tuber;
RX   MEDLINE=92079917; PubMed=1745241;
RA   Kosmann U., Visser R.G.F., Mueller-Roeder B., Willmitzer L.,
RA   Somewald U.;
RT   "Cloning and expression analysis of a potato cDNA that encodes
RT   branching enzyme: evidence for co-expression of starch biosynthetic

```

```

RT   genes.";
RT   Mol. Gen. Genet. 230:39-44 (1991).
CC      -1- FUNCTION: Catalyzes the formation of the alpha-1,6-glucosidic
CC      linkages in glycogen by scission of a 1,4-alpha-linked
CC      oligosaccharide from growing alpha-1,4-glycan chains and the
CC      subsequent attachment of the oligosaccharide to the alpha-1,6
CC      position.
CC      -1- CATALYTIC ACTIVITY: Formation of 1,6-glucosidic linkages of
CC      starch.
CC      -1- PATHWAY: Starch biosynthesis; third step.
CC      -1- SUBUNIT: Monomer.
CC      -1- SUBCELLULAR LOCATION: Chloroplast; amyloplast.
CC      -1- SIMILARITY: Belongs to family 13 of glycosyl hydrolases.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL; X69805; CAA49463.1; -.
CC      PIR; S34730; S34730.
CC      InterPro; IPR006047; Alpha_ami1_cat.
CC      InterPro; IPR004193; Glyco_hydro_13N.
CC      Pfam; PF00128; alpha-amy1ase; 1.
CC      Pfam; PF02822; isoamylase_N; 1.
CC      Starch biosynthesis; Transferase; Glycosyltransferase; Amyloplast;
CC      Chloroplast.
CC      ACT_SITE; 320; 320; BY SIMILARITY.
CC      ACT_SITE; 355; 355; BY SIMILARITY.
CC      ACT_SITE; 360; 360; BY SIMILARITY.
CC      ACT_SITE; 427; 427; BY SIMILARITY.
CC      ACT_SITE; 429; 429; BY SIMILARITY.
CC      ACT_SITE; 484; 484; BY SIMILARITY.
CC      ACT_SITE; 553; 553; BY SIMILARITY.
CC      SEQUENCE; 861 AA; 99083 MW; F3D519AC7CF1EEF2 CRC64;
CC
QY      Query Match          54.4%; Score 37; DB 1; Length 861;
QY      Best Local Similarity 58.3%; Pred. No. 76;
QY      Matches 7; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY
Db      2 IRRKTSSEYNE 13
QY      ||| |||
QY      441 INMGFTGNNEY 452
QY
RESULT 14
AID4_YEAST STANDARD; PRT; 329 AA.
ID   AID4_YEAST
AC   C07747;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Putative aryl-alcohol dehydrogenase AID4 (EC 1.1.1.-).
GN   AID4 OR YDL243C.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
OX   (1)
RN   SEQUENCE FROM N.A.
RC   Alt-Moerbe U., Schneider C., Moro M.;
RA   Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
RA   -1- SIMILARITY: Belongs to the aldo/keto reductase 2 family.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/

```

```

CC or send an email to license@lsb-sib.ch).
CC -----
DR EMBL: Z74291; CAA88823.1; -.
DR PIR: S67807; S67807.
DR GERMOnline: 140486; -.
DR SGD: S0002402; AAD4.
DR InterPro: IPR001395; Aldo/ket_red.
DR Pfam: PF00248; aldo_ket_red; 1.
DR ProDom: PD000288; Aldo/Ket_red; 2.
KM Oxidoreductase.
FT ACT_SITE 105 105
SQ SEQUENCE 329 AA; 36977 MW; 62DBDD0C869610E0 CRC64;

```

```

Query Match 52.9%; Score 36; DB 1; Length 329;
Best Local Similarity 46.2%; Pred. No. 41;
Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

```

```

QY 1 LIRKFTSEYNEY 13
DB 53 VIATKFTGDYKXY 65

```

```

RESULT 15
AADE_YEAST
ID_AADE_YEAST STANDARD; PRT; 376 AA.
AC P42884;
DT 01-NOV-1995 (Rel. 32; Created)
DT 01-NOV-1995 (Rel. 32; Last sequence update)
DT 15-JUL-1999 (Rel. 38; Last annotation update)
DE Putative aryl-alcohol dehydrogenase AAD14 (EC 1.1.1.-).
GN AAD14 OR YNL331C OR N0300.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288c / FY1679;
RA van Dyck L., Pascual-Anur A., Goffeau A.;
RL Submitted (DEC-1994) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Obermaier B., Piravandi E., Rinke M.;
RL Submitted (MAY-1996) to the EMBL/Genbank/DBJ databases.
RN [2]
CC -! SIMILARITY: BELONGS TO THE ALDO/KETO REDUCTASE 2 FAMILY. STRONG,
CC TO YEAST AAD6.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.lsb-sib.ch/announce/
CC or send an email to license@lsb-sib.ch).
CC -----
DR EMBL: X83226; CAA58227.1; -.
DR EMBL: Z71607; CAA96264.1; -.
DR PIR: S51335; S51335.
DR GERMOnline: 143337; -.
DR SGD: S0005275; AAD14.
DR InterPro: IPR001395; Aldo/ket_red.
DR Pfam: PF00248; aldo_ket_red; 1.
DR ProDom: PD000288; Aldo/Ket_red; 1.
KM Oxidoreductase.
FT ACT_SITE 151 151
SQ SEQUENCE 376 AA; 41991 MW; 2265406386938313 CRC64;

```

```

Query Match 52.9%; Score 36; DB 1; Length 376;
Best Local Similarity 46.2%; Pred. No. 47;
Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

```

```

QY 1 LIRKFTSEYNEY 13
DB 53 VIATKFTGDYKXY 65

```

DB 99 VIATKFTGDYKXY 111

Search completed: October 5, 2004, 08:01:43  
Job time : 3.16358 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2004, 07:34:11 ; Search time 6.8817 Seconds

(without alignments)  
596.081 Million cell updates/sec

Title: US-09-805-290A-12  
Perfect score: 68  
Sequence: 1 LIRKFTSEYVEY 13

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

1: SP\_ARCHES: \*  
2: SP\_BACTERIA: \*  
3: SP\_FUNGI: \*  
4: SP\_HUMAN: \*  
5: SP\_INVERTEBRATE: \*  
6: SP\_MAMMAL: \*  
7: SP\_MHC: \*  
8: SP\_ORGANELLE: \*  
9: SP\_PHAGE: \*  
10: SP\_PLANT: \*  
11: SP\_PROTOZOA: \*  
12: SP\_VIRUS: \*  
13: SP\_VIRTEBRATE: \*  
14: SP\_UNCLASSIFIED: \*  
15: SP\_VIRUS: \*  
16: SP\_BACTERIAP: \*  
17: SP\_ARCHAEP: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	46	67.6	1517	5 Q9W0L7	Q9W0L7 drosophila
2	45	66.2	1044	16 Q8XIN5	Q8XIN5 drosophila
3	43	63.2	401	5 Q8BXO	Q8BXO drosophila
4	42	61.8	405	5 Q8B3J8	Q8B3J8 drosophila
5	41	60.3	407	5 Q8B3J7	Q8B3J7 drosophila
6	40.5	59.6	413	17 Q980S0	Q980S0 sulfolobus
7	40	58.8	135	10 Q08130	Q08130 manihot esc
8	40	58.8	171	16 Q7V0X5	Q7V0X5 prochloroc
9	40	58.8	273	16 Q8BTF9	Q8BTF9 lactobacill
10	40	58.8	696	5 Q9VUZ6	Q9VUZ6 drosophila
11	40	58.8	826	10 Q81021	Q81021 drosophila
12	40	58.8	852	10 Q08131	Q08131 manihot esc
13	40	58.8	1028	16 Q81EV1	Q81EV1 bacillus ce
14	40	58.8	1789	5 Q81UD3	Q81UD3 plasmodium
15	40	58.8	179	17 Q97W67	Q97W67 sulfolobus

17	39	57.4	185	16 Q8A810	Q8A810 bacteroides
18	39	57.4	280	5 Q00840	Q00840 dictyostell
19	39	57.4	356	5 Q81D60	Q81D60 plasmodium
20	39	57.4	460	12 Q91EY7	Q91EY7 cydia pomon
21	39	57.4	463	16 Q8P23	Q8P23 xanthomonas
22	39	57.4	463	16 Q8PCE1	Q8PCE1 xanthomonas
23	39	57.4	534	16 Q87G09	Q87G09 vibrio para
24	39	57.4	545	16 Q8B3J1	Q8B3J1 xanthomonas
25	39	57.4	1014	2 Q92H7	Q92H7 yersinia ps
26	38	55.9	182	16 Q8D1Q5	Q8D1Q5 synechococ
27	38	55.9	190	16 Q91F67	Q91F67 staphylococ
28	38	55.9	219	16 Q8F608	Q8F608 leptospira
29	38	55.9	225	12 Q8B0X8	Q8B0X8 peanut clum
30	38	55.9	286	2 Q91B33	Q91B33 clostridium
31	38	55.9	289	2 Q51585	Q51585 plectonema
32	38	55.9	337	17 Q980G4	Q980G4 sulfolobus
33	38	55.9	339	16 Q971F6	Q971F6 clostridium
34	38	55.9	671	11 Q99N10	Q99N10 mus musculu
35	38	55.9	671	11 Q8R499	Q8R499 mus musculu
36	38	55.9	685	11 Q99M66	Q99M66 mus musculu
37	38	55.9	847	10 Q9XIS4	Q9XIS4 phaseolus v
38	38	55.9	877	16 Q88AW5	Q88AW5 pseudomonas
39	38	55.9	2025	12 Q8J7Z4	Q8J7Z4 lumpy skin
40	38	55.9	2025	12 Q91MNO	Q91MNO lumpy skin
41	37	54.4	120	11 Q9CY23	Q9CY23 mus musculu
42	37	54.4	139	2 Q9AMN3	Q9AMN3 clostridium
43	37	54.4	143	16 Q97TF9	Q97TF9 clostridium
44	37	54.4	161	4 Q9H2H8	Q9H2H8 homo sapien
45	37	54.4	161	4 Q961A9	Q961A9 homo sapien

#### ALIGNMENTS

RESULT 1  
ID Q9W0L7 PRELIMINARY; PRT; 1517 AA.  
AC Q9W0L7;  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)  
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
DE CG32479 protein (LID28815p).  
GN CG32479 OR CG12503 OR CG13903.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN=Berkley;  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celiker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman U.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abriil J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
RA Ballew R.M., Baer A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borrova D., Botchan M.R., Bouck J., Brckstein P., Brothier P.,  
RA Butris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.N.,  
RA Dodson K., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrera S., Fleischmann W.,  
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ileguam C.,  
RA Jaisli W., Kalush F., Karpen G.H., Ke Z., Kemison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

```

RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Maitel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Mikhlov G., Mishina N.V., Moberly C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazolo M., Pittman K.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skuski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wastaman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195 (2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Centliker S.E., Adams M.D., Krommiller B., Wan K.H., Holt R.A.,
RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA Banson J., An H., Baldwin D., Banson J., Beeson K.Y., Blum D.A.,
RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek P., Fartan D.,
RA Ferreira M., Frise E., Galle R.F., Garg N.S., George R.A.,
RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA Ibegwam C., Jalali M., Kruse D., Li P., Maitel B., Moshrefi A.,
RA Murtoson T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nuno J.,
RA Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,
RA Phoonanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA Stapleton M., Strong R., Swirskas R., Tector C., Tyler D.,
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.,
RT "Sequencing of Drosophila melanogaster genome."
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Mitra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,
RA Tuzy J.L., Bergman C., Bernan B., Carlson J.W., Centliker S.E.,
RA Clump M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA Krommiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA Seattle S.M., Smith E., Snu S., Smutnick F., Whitfield E.,
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.,
RT "Annotation of Drosophila melanogaster genome."
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Adams M.D., Centliker S.E., Gibbs R.A., Rubin G.M., Venter C.J.,
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RA Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE FROM N.A.
RA STRAIN=Berkeley;
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
RA George R., Gonzalez M., Guarin H., Krommiller B., Li P., Liao G.,
RA Miranda A., Mungall C.J., Nuno J., Pacleb J., Paragas V., Park S.,
RA Patel S., Phuanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA Centliker S.,
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
RA EMBL; AEO03469; AAF47426.2; -
RA EMBL; AY122170; AAMS2682.1; -
RA FlyBase; FBgn0052479; CG32479.
DR GO; GO:0004197; F:cytoskeleton-type endopeptidase activity; IEA.
DR GO; GO:0004221; P:ubiquitin-thiolesterase activity; IEA.
DR GO; GO:0006511; P:ubiquitin-dependent proteolysis; IEA.
DR InterPro; IPR001394; Peptidase_C19.
DR Pfam; PF00443; UCH; 1.
DR PROSITE; PS00973; UCH_2_2; 1.
DR PROSITE; PS0235; UCH_2_3; 1.

```

---

```

SQ SEQUENCE 1517 AA; 16463 MW; 03DB24E4D2A7A77C CRC64;
Query Match 67.6%; Score 46; DB 5; Length 1517;
Best Local Similarity 58.3%; Pred. No. 18;
Matches 7; Conservative 4; Mismatches 1; Indels 0; Gaps 0;
Cy 2 IRRKPTSEVY 13
Db 693 VRRKISSEVNH 704
RESULT 2
ID 08XINS PRELIMINARY; PRT; 1044 AA.
ID 08XINS;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Alpha-mannosidase.
GN CPE2080.
OS Clostridium perfringens.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OC NCBI_TaxID=1502;
OX [1]
RP SEQUENCE FROM N.A.
RA STRAIN=13 / TYPE A;
RA MEDLINE=21664373; PubMed=11792842;
RA Shimizu T., Ohnari K., Hirakawa K., Ohshima K., Yamashita A.,
RA Shibata T., Ogasawara N., Hattori M., Kohara S., Hayashi H.,
RT "Complete genome sequence of Clostridium perfringens, an anaerobic
RT flesh-eater."
RL Proc. Natl. Acad. Sci. U.S.A. 99:996-1001 (2002).
DR EMBL; AP003192; BAB81766.1; -
DR GO; GO:0004559; P:alpha-mannosidase activity; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR006062; Glyco_hydro_38.
DR Pfam; PF01074; Glyco_hydro_38; 1.
KW Complete proteome.
SQ SEQUENCE 1044 AA; 121431 MW; 3CABB79447D42B6F CRC64;
Query Match 66.2%; Score 45; DB 16; Length 1044;
Best Local Similarity 72.7%; Pred. No. 19;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
Cy 2 IRRKPTSEVY 12
Db 201 IRRPPTKEYNE 211
RESULT 3
ID 086BK0 PRELIMINARY; PRT; 401 AA.
AC 086BK0;
DT 01-JUN-2003 (TREMBlrel. 24, Created)
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE CG9614-PJ.
GN PIP OR CG9614.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Megaloptera; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RP [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Centliker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wotman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,

```

RA Abriil J.F., Agbayani A., An H.J., Andrews-Pfannkuch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Bencos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Borkan M.R., Bouck J., Brokstein P., Brothier P.,  
 RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Chertov J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Dudbin K.J., Evangelista C.C., Ferraz C., Fextera S., Fleischmann W.,  
 RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,  
 RA Jallali M., Kalush F., Karpen G.H., Ke Z., Kenison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasako P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusken D.R., Pacleb J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Furl V., Reese M.G.,  
 RA Reiter K., Remington K., Saunders R.D., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spletter E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svitskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.Y., Massaman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodger, Morley K.C., Wu D., Yang S., Yao Q.A., Ye J.,  
 RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,  
 RA "The genome sequence of *Drosophila melanogaster*,"  
 RA Science 287:2185-2195(2000).  
 RA [2]  
 RA SEQUENCE FROM N.A.  
 RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,  
 RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,  
 RA Tupy J.L., Bergman C.M., Berman B.P., Carlson J.W., Gelinker S.E.,  
 RA Clamp M.E., Drysdale R.A., Emmert D., Frise E., de Grey A.D.N.J.,  
 RA Harris N.L., Krommiller B., Marshall B., Milburn G.H., Richter J.,  
 RA Russo S., Searle S.M.J., Smith E., Shu S., Smutnack F.,  
 RA Whitfield E.J., Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J.,  
 RA Lewis S.E.,  
 RA "Annotation of *Drosophila melanogaster* genome,"  
 RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RA [3]  
 RA SEQUENCE FROM N.A.  
 RA Flybase;  
 RA Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
 RA [4]  
 RA SEQUENCE FROM N.A.  
 RA Flybase;  
 RA Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.  
 RA EMBL, AE003517, AAC01227.1, -  
 RA InterPro, IPR007734, HS2ST.  
 RA Pfam, PF05040, HS2ST, 1.  
 RA SEQUENCE 401 AA; 47706 MW; EBA62FF4F7338B CRC64;  
 RA Lewis S.E.;  
 RA "Annotation of *Drosophila melanogaster* genome,"  
 RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RA [3]  
 RA SEQUENCE FROM N.A.  
 RA Flybase;  
 RA Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
 RA [4]  
 RA SEQUENCE FROM N.A.  
 RA Flybase;  
 RA Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.  
 RA EMBL, AE003517, AAC01227.1, -  
 RA InterPro, IPR007734, HS2ST.  
 RA Pfam, PF05040, HS2ST, 1.  
 RA SEQUENCE 405 AA; 47803 MW; 5807FC09A5CA264C CRC64;

DE CG9614-PH.  
 GN PIP OR CG9614.  
 OS *Drosophila melanogaster* (fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephyrididae; Drosophilidae; Drosophila.  
 OK NCBI\_taxid=7227;  
 RN  
 RP SEQUENCE FROM N.A.  
 RP MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.H., Blazer V.G., Champs M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gaber G.L.,  
 RA Abriil J.F., Agbayani A., An H.J., Andrews-Pfannkuch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Bencos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Borkan M.R., Bouck J., Brokstein P., Brothier P.,  
 RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Chertov J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Dudbin K.J., Evangelista C.C., Ferraz C., Fextera S., Fleischmann W.,  
 RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,  
 RA Jallali M., Kalush F., Karpen G.H., Ke Z., Kenison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasako P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusken D.R., Pacleb J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Furl V., Reese M.G.,  
 RA Reiter K., Remington K., Saunders R.D., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spletter E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svitskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.Y., Massaman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodger, Morley K.C., Wu D., Yang S., Yao Q.A., Ye J.,  
 RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,  
 RA "The genome sequence of *Drosophila melanogaster*,"  
 RA Science 287:2185-2195(2000).  
 RA [2]  
 RA SEQUENCE FROM N.A.  
 RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,  
 RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,  
 RA Tupy J.L., Bergman C.M., Berman B.P., Carlson J.W., Gelinker S.E.,  
 RA Clamp M.E., Drysdale R.A., Emmert D., Frise E., de Grey A.D.N.J.,  
 RA Harris N.L., Krommiller B., Marshall B., Milburn G.H., Richter J.,  
 RA Russo S., Searle S.M.J., Smith E., Shu S., Smutnack F.,  
 RA Whitfield E.J., Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J.,  
 RA Lewis S.E.;  
 RA "Annotation of *Drosophila melanogaster* genome,"  
 RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RA [3]  
 RA SEQUENCE FROM N.A.  
 RA Flybase;  
 RA Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
 RA [4]  
 RA SEQUENCE FROM N.A.  
 RA Flybase;  
 RA Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.  
 RA EMBL, AE003517, AAC01227.1, -  
 RA InterPro, IPR007734, HS2ST.  
 RA Pfam, PF05040, HS2ST, 1.  
 RA SEQUENCE 405 AA; 47803 MW; 5807FC09A5CA264C CRC64;

```

Query Match          61.8%; Score 42; DB 5; Length 405;
Best Local Similarity 53.8%; Pred. No. 24;
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY      1 LIRKFTSYNEY 13
DB      368 MYRNFTHEYDFY 380

RESULT 5
Q86B7 PRELIMINARY; PRT; 407 AA.
AC Q86B7;
DT 01-JUN-2003 (TREMBlrel. 24, Created)
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE CG9614-PG.
GN PIP OR CG9614.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amaralides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Burton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H., Blazer R.G., Champe M., Peiffer B.D.,
RA Abill U.F., Agbayani A., An H.J., Andrews-Pfankuch C., Baldwin D.,
RA Ballew R.M., Basu A., Bakendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Butts J.M., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Dublin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibsen C.,
RA Jatala W., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kuip D., Lai Z.,
RA Liakou P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Matvei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclab J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spraker E., Spreading A.C., Stapleton M., Strong R., Sun E.,
RA Svrstkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.Y., Wasserman D.A., Weinstein G.M., Weisenbach J.,
RA Williams S.M., Wocagel, Morley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA Yen R.F., Zaveri U.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RN Science 287:2185-2195(2000).
RP SEQUENCE FROM N.A.
RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,
RA Tupy J.L., Bergman C.M., Berman B.P., Carlson J.W., Celinker S.E.,
RA Clamp M.E., Drysdale R.A., Emmert D., Frise E., de Grey A.D.N.J.,
RA Harris N.L., Krommiller B., Marshall B., Millburn G.H., Richter J.,
RA Russo S., Searle S.M.J., Smith E., Shu S., Smutnick F.,
RA Whitfield E.J., Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J.,

```

```

RA Lewis S.E.;
RT "Annotation of Drosophila melanogaster genome.";
RN Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA FlyBase;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA FlyBase;
RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE003517; AAC1228.1;
DR InterPro; IPR007734; HS2ST.
DR Pfam; PF05040; HS2ST; 1.
SQ SEQUENCE 407 AA; 48213 MW; B752EFE00F98C3AC CRC64;

Query Match          60.3%; Score 41; DB 5; Length 407;
Best Local Similarity 53.8%; Pred. No. 18;
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY      1 LIRKFTSYNEY 13
DB      369 MIRKFTNEYDFY 381

RESULT 6
Q980S0 PRELIMINARY; PRT; 413 AA.
AC Q980S0;
DT 01-OCT-2001 (TREMBlrel. 18, Created)
DT 01-OCT-2001 (TREMBlrel. 18, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE DNA/pantothenate metabolism flavoprotein (dtp).
GN DTP OR SSO0210.
OS Sulfolobus solfataricus.
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC Sulfolobus.
OX NCBI_TaxID=2287;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=ATCC 35092 / DSM 1617 / P2;
RA MEDLINE=21332296; PubMed=11427726;
RA Shew Q., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,
RA Aweyer M.J., Chan-Weiher C.C.-Y., Clausen I.G., Curtis B.A.,
RA De Moors A., Brauso G., Fletcher C., Gordon P.M.K.,
RA Heikamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,
RA Thi-Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,
RA Charlebois R.L., Doolittle W.F., Duguet M., Gaasterland T.,
RA Garrett R.A., Ragan M.A., Sensen C.W., Van der Oost J.;
RT "The complete genome of the crenarchaeon Sulfolobus solfataricus P2.";
RN Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).
RP SEQUENCE FROM N.A.
RX PIR; B90162; B90162.
RA EMBL; AE006658; AAK40553.1;
RT "The complete genome of the crenarchaeon Sulfolobus solfataricus P2.";
RN Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).
RP SEQUENCE FROM N.A.
RX InterPro; IPR007085; DTP_C.
RA InterPro; IPR003182; Flavoprotein.
DR Pfam; PF04127; dtp; 1.
DR Pfam; PF02441; Flavoprotein; 1.
DR TIGRfam; TIGR00521; coaBC_dtp; 1.
KV Complete proteome.
SQ SEQUENCE 413 AA; 45864 MW; 79569D38D49A5219 CRC64;

Query Match          59.6%; Score 40.5; DB 17; Length 413;
Best Local Similarity 64.3%; Pred. No. 47;
Matches 9; Conservative 2; Mismatches 0; Indels 3; Gaps 1;

QY      2 IRRK---FTSEYNE 12
DB      365 VRRKDTFSSSYNE 378

RESULT 7
Q008130

```

ID Q08130 PRELIMINARY; PRT: 155 AA.  
 AC Q08130;  
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
 DE Branching enzyme (Fragment).  
 OS Manihot esculenta (Cassava) (Manioc).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
 OC eurosids I; Malpighiales; Euphorbiaceae; Crotonoideae; Manihoteae;  
 OC Manihot.  
 NC NCB1\_TaxID=3983;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93099233; PubMed=1281436;  
 RA Salehuzzaman S.N., Jacobsen E., Vlasser R.G.,  
 RT "Cloning, partial sequencing and expression of a cDNA coding for  
 RT branching enzyme in cassava."  
 RL Plant Mol. Biol. 20:809-819(1992).  
 DR EMBL; X69723; CAA49371.1; -.  
 FT NON\_TER 1 155  
 FT NON\_TER 1 155  
 SQ SEQUENCE 155 AA; 17720 MW; 67EC3180023A5052 CRC64;

Query Match 58.8%; Score 40; DB 10; Length 155;  
 Best Local Similarity 58.3%; Pred. No. 21;  
 Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 IIRKFTSEYNEV 13  
 Db 24 INNAFTGDYNEY 35

RESULT 8  
 ID Q7V0X5 PRELIMINARY; PRT: 171 AA.  
 AC Q7V0X5;  
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Phosphoribosyl transferase (EC 2.4.2.7).  
 GN APT OR PMM112.  
 OS Prochlorococcus marinus subsp. pastoris (strain CCMP 1378 / MED4).  
 OC Bacteria; Cyanobacteria; Prochlorophytes; Prochlorococcaceae;  
 OC Prochlorococcus.  
 NC NCB1\_TaxID=59919;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=2282696; PubMed=12917642;  
 RA Rocap G., Larimer F.W., Lamerdin J., Malfatti S., Chain P.,  
 RA Algren N.A., Arellano A., Coleman U., Hauser L., Hess W.R.,  
 RA Johnson Z.I., Land M., Lindell D., Post A.F., Regala W., Shah M.,  
 RA Shaw S.L., Steglich C., Sullivan M.B., Ting C.S., Tolonen A.,  
 RA Webb E.A., Zinser B.R., Chisholm S.W.;  
 RT "Genomic divergence in two Prochlorococcus ecotypes reflects oceanic  
 RT niche differentiation."  
 RL Nature 424:1042-1047(2003).  
 DR EMBL; BX572093; CAE19581.1; -.  
 KW Glycoyltransferase; Transferase; Complete proteome.  
 SQ SEQUENCE 171 AA; 19114 MW; 1615FD8351848565 CRC64;

Query Match 58.8%; Score 40; DB 16; Length 171;  
 Best Local Similarity 58.3%; Pred. No. 23;  
 Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 IIRKFTSEYNEV 12  
 Db 88 LIKKYTLLEYGE 99

RESULT 9  
 Q081F9 PRELIMINARY; PRT: 273 AA.  
 ID Q081F9

AC Q081F9;  
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Transcription regulator.  
 GN LP\_3006.  
 OS Lactobacillus plantarum.  
 OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;  
 OC Lactobacillus.  
 NC NCB1\_TaxID=1590;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NCIMB 8826 / WCPs1.  
 RX MEDLINE=22480296; PubMed=12566566;  
 RA Kleerebezem M., Boekhorst J., van Kranenburg R., Molenaar D.,  
 RA Kuipers O.P., Leer R., Tarchini R., Peters S.A., Sandbrink H.M.,  
 RA Piers M.W.E.J., Stiekema W., Klein Lankhorst R.M., Bron P.A.,  
 RA Hofer S.M., Nierop Groot M.N., Kerkhoven R., De Vries M., Ursing B.,  
 RA De Vos W.M., Stezen R.J.;  
 RT "Complete genome sequence of Lactobacillus plantarum WCPs1."  
 RL Proc. Natl. Acad. Sci. U.S.A. 100:1990-1995(2003).  
 DR EMBL; AL935260; CAD65187.1; -.  
 DR GO: GO:0003677; F:DNA binding; IEA.  
 DR InterPro; IPR001387; HTH\_3.  
 DR Pfam; PF01381; HTH\_3; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 273 AA; 30562 MW; A1F2A3CB88144F3A CRC64;

Query Match 58.8%; Score 40; DB 16; Length 273;  
 Best Local Similarity 63.6%; Pred. No. 38;  
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 IIRKFTSEYNEV 11  
 Db 179 IIRKFTGKYN 189

RESULT 10  
 ID Q9VUZ6 PRELIMINARY; PRT: 696 AA.  
 AC Q9VUZ6;  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE CG5018 protein.  
 GN CG5018.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 NC NCB1\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Berkeley;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amandis P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang O., Chen L.X.,  
 RA Brandon R.C., Rogers Y.H.C., Blakey R.G., Champé M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abiri J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballaw R.M., Basu A., Baxendale J., Bayraktaroglu U., Beasley E.M.,  
 RA Beeson K.Y., Bencos P.V., Bertan B.P., Bhandari D., Bolshakov S.,  
 RA Borkov D., Botchan M.R., Bouck J., Brockstein P., Brothier P.,  
 RA Burtis K.C., Busan D.A., Butler H., Cadden E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dantke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Foster C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glasser K.,  
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,

```

RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Kerchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mateja B., McInosh T.C., McLeod M.P., McPherson D.,
RA Merulov G., Milstina N.V., Mobarry C., Morris J., Koshirei A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spredling A.C., Stapleton M., Strong R., Sun E.,
RA Switskes R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong M., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
RA "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
DR EMBL, AE003528; AAF49528.1; -.
DR FlyBase; FBgn0036578; CG5018.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 6.
DR SMART; SMO0320; WD40; 5.
DR PROSITE; PSS0294; WD_REPEATS_REGION; 2.
KW Repeat; WD repeat.
SQ SEQUENCE 696 AA; 78667 MW; 6654706B927D7D26 CRC64;

Query Match 58.8%; Score 40; DB 5; Length 696;
Best Local Similarity 61.5%; Pred. No. 1e+02;
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Cy 1 LIRKFTSEVNEY 13
Db 567 LVNRIFTCEINYE 579

RESULT 11
Q8T021 PRELIMINARY; PRT; 696 AA.
ID Q8T021;
AC Q8T021;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE LD39110P.
GN CG5018.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley.
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA Munoz J., Paolel J., Paragas V., Park S., Phoumenavong S., Wan K.,
RA Yu C., Lewis S.E., Rubin G.M., Celnik S.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY069620; AAJ39765.1; -.
DR FlyBase; FBgn0036578; CG5018.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 6.
DR SMART; SMO0320; WD40; 7.
DR PROSITE; PSS0294; WD_REPEATS_REGION; 2.
KW Repeat; WD repeat.
SQ SEQUENCE 696 AA; 78695 MW; 1A932C915A2E88E1 CRC64;

Query Match 58.8%; Score 40; DB 5; Length 696;
Best Local Similarity 61.5%; Pred. No. 1e+02;
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

```

```

Cy 1 LIRKFTSEVNEY 13
Db 567 LVNRIFTCEINYE 579

RESULT 12
Q41059 PRELIMINARY; PRT; 826 AA.
ID Q41059;
AC Q41059;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Starch branching enzyme II precursor (Fragment).
GN SBEII.
OS Pisum sativum (Garden pea).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosid 1; Fabales; Fabaceae; Papilionoideae; Viciae; Pisum.
OX NCBI_TaxID=3888;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95201826; PubMed=7894509;
RA Burton R.A., Bewley J.D., Smith A.M., Bhattacharya M.K., Tatge H.,
RA King S., Bull V., Hamilton W.D.O., Martin C.;
RT "Starch branching enzymes belonging to distinct enzyme families are
RL Plant J. 7:3-15(1995).
DR EMBL; X80010; CAA56320.1; -.
DR PIR; T06494; T06494.
DR GO; GO:0004555; F:alpha-amylase activity; IEA.
DR GO; GO:0004553; F:hydrolyase activity; hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR006047; Alpha-amyl_cat.
DR InterPro; IPR004193; Glyco_hydro_13N.
DR Pfam; PF00128; alpha-amylase; 1.
DR Pfam; PF02922; isomylase_N; 1.
KW TRANSIT peptide.
FT NON_TER 1
FT TRANSIT <1 58 POTENTIAL.
FT CHAIN 59 826 POTENTIAL.
SQ SEQUENCE 826 AA; 94191 MW; 46B9D165C8E9C9C CRC64;

Query Match 58.8%; Score 40; DB 10; Length 826;
Best Local Similarity 58.3%; Pred. No. 1.2e+02;
Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Cy 2 IIRKFTSEVNEY 13
Db 421 INMAFTGDYNEY 432

RESULT 13
Q08131 PRELIMINARY; PRT; 852 AA.
ID Q08131;
AC Q08131;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE 1,4-alpha-glucan branching enzyme precursor (EC 2.4.1.18) (1,4-alpha-
DE glucan branching enzyme R-3) (Starch branching enzyme) (Q-enzyme).
GN SBE1 OR SBE.
OS Manihot esculenta (Cassava) (Manioc).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosid 1; Malpighiales; Euphorbiaceae; Crotonoideae; Manihoteae;
OC Manihot.
OX NCBI_TaxID=3983;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. M.COL. 22; TISSUE=tuberous root;
RA Salehuzzaman S.M., Jacobsen E., Visser R.G.F.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.

```



RN [2]  
 RP SEQUENCE OF 661-765 FROM N.A.  
 RX MEDLINE=93092923; PubMed=1281436;  
 RA Salehuzzaman S.N., Jacobsen E., Vissers R.G.;  
 RT "Cloning, partial sequencing and expression of a cDNA coding for  
 branching enzyme in cassava."  
 RL Plant Mol. Biol. 20:809-819(1992).  
 CC -1- CATALYTIC ACTIVITY: FORMATION OF 1,6-GLUCOSIDIC LINKAGES OF  
 GLYCOGEN.  
 CC -1- PATHWAY: THIRD STEP IN STARCH BIOSYNTHESIS.  
 CC -1- SUBUNIT: MONOMER.  
 CC -1- SUBCELLULAR LOCATION: AMYLOPLAST.  
 CC -1- SIMILARITY: BELONGS TO FAMILY 13 OF GLYCOSYL HYDROLASES, ALSO  
 KNOWN AS THE ALPHA-AMYLASE FAMILY.  
 CC EMBL: X77012; CAA54308.1; -  
 DR EMBL: X69714; CAA49372.1; -  
 DR GO: GO:0003844; F.1.4-alpha-glucan branching enzyme activity; IEA.  
 DR GO: GO:0004556; F.alpha-amyase activity; IEA.  
 DR GO: GO:0004553; F.hydrolase activity, hydrolyzing O-glycosyl . . . ; IEA.  
 DR GO: GO:0016577; F.transferase activity, transferring glycosyl . . . ; IEA.  
 DR GO: GO:0005875; P.carbohydrate metabolism; IEA.  
 DR GO: GO:0019252; P.starch biosynthesis; IEA.  
 DR InterPro: IPR006047; Alpha\_amy1\_cat.  
 DR InterPro: IPR004193; Glyco\_Hydro\_13N.  
 DR Pfam: PF00128; alpha-amyase; 1.  
 DR Pfam: PF02922; isoamylase N; 1.  
 DR GlycoStarch: Starch biosynthesis; Transferase;  
 KM Transic peptide.  
 FT TRANSIT 1 74 POTENTIAL.  
 FT CHAIN 1 852 1,4-ALPHA-GLUCAN BRANCHING ENZYME.  
 SQ SEQUENCE 852 AA; 96684 MW; 1FBAE63B674C65 CRC64;  
 QY  
 Db 2 IRRKFTSEYNEY 13  
 438 INVAFTGDYNEY 449  
 Query Match 58.8%; Score 40; DB 10; Length 852;  
 Best Local Similarity 58.3%; Pred. No. 1.3e+02;  
 Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Query Match 58.8%; Score 40; DB 16; Length 1028;  
 Best Local Similarity 58.9%; Pred. No. 1.5e+02;  
 Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 5 KFTSEYNEY 13  
 Db 307 KFTSEYNEY 315  
 RESULT 15  
 ID 081UT3 PRELIMINARY; PRT; 1789 AA.  
 AC 081UT3;  
 DT 01-MAR-2003 (TREMBlrel. 23, Created)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
 DE Hypothetical protein.  
 GN PF10\_0099.  
 OS Plasmodium falciparum (isolate 3D7).  
 CC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.  
 CC NCBI\_Taxid=36329;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=3D7;  
 RX MEDLINE=22255705; PubMed=12368864;  
 RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,  
 RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,  
 RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,  
 RA Chan M.-S., Nene V., Shallow S.J., Sub B., Peterson J., Anguioni S.,  
 RA Perera M., Allen J., Fairlamb A.H., Frumholz M.J., Ross D.S., Ralph S.A.,  
 RA Martin D.M.A., Fairlamb A.H., Sudraman G.M., Kung'u G.,  
 RA McFadden G.I., Cummings L.M., Subramanian G.M., Kung'u G.,  
 RA Venker J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,  
 RA Fraser C.M., Barrall B.;  
 RT "Genome sequence of the human malaria parasite Plasmodium  
 falciparum."  
 RL Nature 419:498-511(2002).  
 DR EMBL: AB014830; AAN35297.1; -  
 KW Hypothetical protein.  
 SQ SEQUENCE 1789 AA; 211507 MW; 30C6F9F39D13A7C7 CRC64;  
 QY 1 IRRKFTSEYNEY 13  
 Db 633 ILRNFFFSYNYKY 645  
 Search completed: October 5, 2004, 08:13:22  
 Job time : 10.8812 secs

This Page Blank (uspioj)



XX SQ Sequence 13 AA; 100.0%; Score 68; DB 4; Length 13;  
 Query Match Best Local Similarity 100.0%; Pred. NO. 4.7e-05;  
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LIRRKFTSEYNEY 13  
 DB 1 LIRRKFTSEYNEY 13

RESULT 2  
 AAE10556  
 ID AAE10556 standard; peptide; 130 AA.  
 AC AAE10556;  
 DT 10-DEC-2001 (first entry)  
 DE HPL inhibiting VHH fragment, HPL #18 from llama species.  
 XX  
 KW Llama antibody; camelid; anorectic; heavy chain variable domain; VHH;  
 KW human dietary enzyme inhibitor; medicament; human pancreatic lipase; HPL;  
 KW food; human gastric lipase; HGL; cosmetic control; body weight.  
 XX  
 OS Lama sp.  
 FH Key Location/Qualifiers  
 FT Region 31..35  
 FT /label= CDR1  
 FT /note= "Complementarity determining region 1"  
 FT 50..64  
 FT /label= CDR2  
 FT /note= "Complementarity determining region 2"  
 FT 98..110  
 FT /label= CDR3  
 FT /note= "Complementarity determining region 3"

EP1134231-A1.  
 PD 19-SEP-2001.  
 PF 20-FEB-2001; 2001EP-00200703.  
 XX  
 PR 14-MAR-2000; 2000EP-00200930.  
 XX  
 PA (UNIL ) UNILEVER NV.  
 PA (UNIL ) UNILEVER PLC.  
 XX  
 PI Bezemer S, Van De Burg M, De Haard JWM, Tareilus E;  
 DR WPI; 2001-572718/65.  
 XX  
 PT New antibody or its fragments for inhibiting human dietary enzymes,  
 FT useful for cosmetic control of body weight of human beings, comprises  
 PT heavy chain variable domain derived from immunoglobulin naturally devoid  
 of light chains.  
 XX  
 PS Example 2; Page 10; 37pp; English.  
 CC The patent discloses antibodies or their fragments comprising a heavy  
 CC chain variable domain (VHH) derived from an immunoglobulin naturally  
 CC devoid of light chains specific for inhibiting human dietary enzymes. The  
 CC antibodies of the invention are useful for the preparation of medicaments  
 CC or food for inhibiting the activity of one or more human dietary enzymes  
 CC especially human pancreatic lipase (HPL) or human gastric lipase (HGL)  
 CC which are useful for the cosmetic control of body weight of human beings.  
 CC The present peptide sequence is HPL inhibiting VHH fragment, HPL #18 from  
 CC llama (camelid) species  
 XX  
 SQ Sequence 130 AA;

Query Match 100.0%; Score 68; DB 4; Length 130;  
 Best Local Similarity 100.0%; Pred. NO. 0.00052;  
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LIRRKFTSEYNEY 13  
 DB 98 LIRRKFTSEYNEY 110

RESULT 3  
 ABB68762  
 ID ABB68762 standard; protein; 1003 AA.  
 AC ABB68762;  
 DT 26-MAR-2002 (first entry)  
 DE Drosophila melanogaster polypeptide SEQ ID NO 33078.  
 XX  
 KW Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical.  
 XX  
 OS Drosophila melanogaster.  
 XX  
 PN WO200171042-A2.  
 XX  
 PD 27-SEP-2001.  
 PF 23-MAR-2001; 2001WO-US009231.  
 XX  
 PR 23-MAR-2000; 2000US-0191637P.  
 PR 11-JUL-2000; 2000US-00614150.  
 XX  
 PA (PEKE ) PE CORP NY.  
 XX  
 PI Venter JC, Adams M, Li PWD, Myers EW;  
 XX  
 DR WPI; 2001-656860/75.  
 DR N-PSDB; ABL12865.  
 XX  
 PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signaling and cell-cell  
 PT interactions.  
 XX  
 PS Disclosure; SEQ ID NO 33078; 21pp + Sequence Listing; English.  
 XX  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL16176-ABL130511), expressed DNA  
 CC sequences (AB372072) and the encoded proteins (AB57737-  
 CC AB372072). The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 1003 AA;

Query Match 67.6%; Score 46; DB 4; Length 1003;  
 Best Local Similarity 58.3%; Pred. NO. 30;  
 Matches 7; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

OY 2 LIRRKFTSEYNEY 13  
 DB 179 VRRKXSEYNNH 190

RESULT 4  
 AAE10550  
 ID AAE10550 standard; peptide; 13 AA.  
 AC AAE10550;

XX 10-DEC-2001 (first entry)  
 DT Llama species antibody VH CDR3 #19.  
 XX  
 DE Llama species antibody VH CDR3 #19.  
 XX  
 KW Llama antibody; camelid; anorectic; heavy chain variable domain; VH;  
 KW human dietary enzyme inhibitor; medicament; human pancreatic lipase; HPL;  
 KW food; human gastric lipase; HGL; cosmetic control; body weight;  
 KW complementarity determining region 3; CDR3.  
 XX  
 OS Lama sp.  
 XX  
 PN EP1134231-A1.  
 XX  
 PD 19-SEP-2001.  
 XX  
 PF 20-FEB-2001; 2001EP-00200703.  
 XX  
 PR 14-MAR-2000; 2000EP-00200930.  
 XX  
 PA (UNIL ) UNILEVER NV.  
 PA (UNIL ) UNILEVER PLC.  
 XX  
 PI Bezemer S, Van De Burg M, De Haard JMW, Tarelius E;  
 XX  
 DR WPI; 2001-572718/65.  
 XX  
 PT New antibody or its fragments for inhibiting human dietary enzymes,  
 PT useful for cosmetic control of body weight of human beings, comprises  
 PT heavy chain variable domain derived from immunoglobulin naturally devoid  
 PT of light chains.  
 XX  
 PS Claim 4; Page 29; 37pp; English.  
 XX  
 CC The patent discloses antibodies or their fragments comprising a heavy  
 CC chain variable domain (VH) derived from an immunoglobulin naturally  
 CC devoid of light chains specific for inhibiting human dietary enzymes. The  
 CC antibodies of the invention are useful for the preparation of medicaments  
 CC or food for inhibiting the activity of one or more human dietary enzymes  
 CC especially human pancreatic lipase (HPL) or human gastric lipase (HGL).  
 CC which are useful for the cosmetic control of body weight of human beings.  
 CC The present peptide sequence is a complementarity determining region 3  
 CC (CDR3) of llama species (camelid) antibody VH region  
 XX  
 SQ Sequence 13 AA;  
 XX  
 Query Match 60.3%; Score 41; DB 4; Length 13;  
 Best Local Similarity 58.3%; Pred. No. 2.4;  
 Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;  
 OY 2 IRRKFTSEYNEY 13  
 ||:|||||:  
 2 IRRKFTSGYSHY 13  
 Db  
 RESULT 5  
 ABB61188  
 ID ABB61188 standard; protein; 696 AA.  
 XX  
 AC ABB61188;  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE Drosophila melanogaster polypeptide SEQ ID NO 10356.  
 XX  
 KW Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical.  
 XX  
 OS Drosophila melanogaster.  
 XX  
 PN WO200171042-A2.  
 XX  
 PD 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US009231.  
 PF  
 XX 23-MAR-2000; 2000US-0191637P.  
 PR  
 PR 11-JUL-2000; 2000US-00614150.  
 XX  
 PA (PEKE ) PE CORP NY.  
 XX  
 PI Venter JC, Adams M, Li PWD, Myers EW;  
 XX  
 DR WPI; 2001-656860/75.  
 DR N-PSDB; ABL05291.  
 XX  
 PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions.  
 XX  
 PS Disclosure; SEQ ID NO 10356; 21pp + Sequence Listing; English.  
 XX  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 CC sequences (ABL01840-ABL16175) and the encoded proteins (AB857737-  
 CC AB872072). The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pcr\_sequences  
 XX  
 SQ Sequence 696 AA;  
 XX  
 Query Match 58.8%; Score 40; DB 4; Length 696;  
 Best Local Similarity 61.5%; Pred. No. 2.3e+02;  
 Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;  
 OY 1 IRRKFTSEYNEY 13  
 ||:|||||:  
 567 LVNRFTCEYNEY 579  
 Db  
 RESULT 6  
 ABBU4503  
 ID ABBU4503 standard; protein; 293 AA.  
 XX  
 AC ABBU4503;  
 XX  
 DT 19-JUN-2003 (first entry)  
 XX  
 DE Protein encoded by prokaryotic essential gene #30030.  
 XX  
 KW Antisense; prokaryotic essential gene; cell proliferation; drug design.  
 XX  
 OS Streptococcus mutans.  
 XX  
 PN WO200277183-A2.  
 XX  
 PD 03-OCT-2002.  
 XX  
 PF 21-MAR-2002; 2002WO-US009107.  
 XX  
 PR 21-MAR-2001; 2001US-00815242.  
 PR 06-SEP-2001; 2001US-00948993.  
 PR 25-OCT-2001; 2001US-0342823P.  
 PR 08-FEB-2002; 2002US-00072851.  
 PR 06-MAR-2002; 2002US-0362699P.  
 XX  
 PA (ELIT-) ELITRA PHARM INC.  
 XX  
 PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
 PI Wali D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
 XX  
 DR WPI; 2003-029926/02.

DR N-PSDB; A048373.

XX New antisense nucleic acids, useful for identifying proteins or screening  
PT for homologous nucleic acids required for cellular proliferation to  
XX isolate candidate molecules for rational drug discovery programs.

PS Claim 25; SEQ ID NO 72427; 1766bp; English.

CC The invention relates to an isolated nucleic acid comprising any one of  
CC the 6213 antisense sequences given in the specification where expression  
CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
CC (1) a vector comprising a promoter operably linked to the nucleic acid  
CC encoding a polypeptide whose expression is inhibited by the antisense  
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
CC polypeptide or its fragment whose expression is inhibited by the  
CC antisense nucleic acid; (4) an antibody capable of specifically binding  
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
CC proliferation or the activity of a gene in an operon required for  
CC proliferation; (7) identifying a compound that influences the activity of  
CC the gene product or that has an activity against a biological pathway  
CC required for proliferation, or that inhibits cellular proliferation; (8)  
CC identifying a gene required for cellular proliferation or the biological  
CC pathway in which a proliferation-required gene or its gene product lies  
CC or a gene on which the test compound that inhibits proliferation of an  
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of  
CC strains; or (13) identifying the target of a compound that inhibits the  
CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
CC the target prokaryotic essential genes. Note: The sequence date for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences

CC Sequence 293 AA;

Query Match 57.4%; Score 39; DB 6; Length 293;  
Best Local Similarity 58.3%; Pred. NO. 1.4e+02;  
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 IRRKFTSEYNEY 13  
DB 247 VREKTSUNDY 258

RESULT 7  
AAR82067  
ID AAR82067 standard; protein; 1422 AA.

AC AAR82067;

DT 27-AUG-2003 (revised)  
DT 02-JUL-1996 (first entry)

DE Hepatitis GB virus (HGBV) clone protein prod.

XX Hepatitis GB virus; HGBV; diagnosis; treatment; vaccine; reagents; non-A;  
KM non-B; non-C; non-D; clone; tamarin; infected plasma;  
KM lambda phage; cDNA library.

XX Hepatitis G virus.

XX Key Location/Qualifiers

FT Misc-difference 1..1422  
FT /note="others correspond to degenerate or STOP codons in  
FT AAT00052"

PN W09521922-A2.

XX 17-AUG-1995.

PF 14-FEB-1995; 95WC-US002118.

XX 14-FEB-1994; 94US-00196030.

PR 13-MAY-1994; 94US-00242654.

PR 29-JUL-1994; 94US-00283314.

PR 23-NOV-1994; 94US-00344185.

PR 27-JUN-1995; 95US-00344557.

XX (ABBO) ABBOTT LAB.

PI Simons JN, Pilot-Matias TV, Dawson GU, Schlauder GC, Desai SM;

PI Leary TP, Muehrhoff AS, Erker JC, Buix SL, Mushahwar IK;

DR N-PSDB; AAT00052.

XX WPI: 1995-293123/38.

PT Non-A, non-B, non-C, non-D, non-E Hepatitis virus reagents - useful for

XX diagnosis and therapy of hepatitis GB virus.

PS Example 5; Page 240-244; 661pp; English.

CC Double stranded hepatitis GB virus (HGBV) DNA cbtd. from HGBV infected

CC tamarin plasma, using standard procedures, was used to prepare a lambda

CC phage HGBV cDNA library. The cDNA clone AAT00052, which encodes the

CC protein AAR82066-71 (the 6 possible reading frames), was rescued from

CC the lambda phage, searched against a sequence database and found to be an

CC unique HGBV sequence. Reagents which comprise the HGBV DNA, or its

CC protein prods. can be used for the diagnosis, therapy or in a vaccine to

CC prevent HGBV infection. (Updated on 27-AUG-2003 to correct OS field.)

XX Sequence 1422 AA;

Query Match 57.4%; Score 39; DB 2; Length 1422;  
Best Local Similarity 38.5%; Pred. NO. 7.1e+02;  
Matches 5; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 1 IRRKFTSEYNEY 13  
DB 1222 WVRRLVNDYRNF 1234

RESULT 8  
AAB09037  
ID AAB09037 standard; protein; 1422 AA.

AC AAB09037;

DT 06-AUG-2003 (revised)  
DT 30-AUG-2000 (first entry)

DE Hepatitis GB virus protein sequence SEQ ID NO:82.

XX Hepatitis GB virus; HGBV; diagnosis; therapeutic; immunogenic; infection;

XX detection; characterisation; hepatitis.

XX Hepatitis GB virus.

XX US6051374-A.

XX 18-APR-2000.

XX 07-JUN-1995; 95US-00488445.

XX 14-FEB-1994; 94US-00196030.

XX 13-MAY-1994; 94US-00242654.

XX 29-JUL-1994; 94US-00283314.

XX 23-NOV-1994; 94US-00344185.

XX 23-NOV-1994; 94US-00344190.

XX	30-JAN-1995;	95US-00377557.
XX	(ABBO ) ABBOTT LAB.	
PI	Dawson GU, Leary TP, Muerhoff AS,	Pilot-Matias TJ, Buick SL;
XX	Mushawat IR, Simons JN, Desai SM,	Etker JC, Schlauder GG;
DR	WPI; 2000-338307/29.	
XX		
PT	Detecting target hepatitis GB virus nucleic acid in a test sample	
PT	suspected of containing HGBV comprises reacting the test sample the HGBV	
PT	polynucleotide probe and detecting the complex that contains target HGBV.	
PS	Example 5; Col 215-224; 369pp; English.	
XX		
CC	The present invention describe a method for detecting target hepatitis GB	
CC	virus (HGBV) nucleic acid (THN) in a test sample (T) suspected of	
CC	containing HGBV. The method involves reacting (T) with a HGBV	
CC	polynucleotide probe (I) containing 15 contiguous nucleotides, and which	
CC	selectively hybridizes to the HGBV genome or its full complement, and	
CC	detecting the complex that contains THN, indicating the presence of	
CC	target HGBV. The method is used for detecting target HGBV nucleic acid in	
CC	the test sample suspected of containing HGBV and for characterisation of	
CC	newly ascertained etiological agent of non-A, non-B, non-C, non-D and non	
CC	-E hepatitis causing agents collectively termed as hepatitis GB virus.	
CC	AA55270 to AA55489 and AB08985 to AB09450 represent nucleotide and	
CC	protein sequences used in the exemplification of the present invention.	
CC	(updated on 06-AUG-2003 to correct OS field.)	
XX		
SQ	Sequence 1422 AA:	
	Query Match	57.4%; Score 39; DB 3; Length 1422;
	Best Local Similarity	38.5%; Pred. No. 7.1e+02;
	Matches 5; Conservative	5; Mismatches 3; Indels 0; Gaps 0
OY	1 LIRRKFTSEYNEX 13	
	::::: :::: :	
Db	1222 MKRRLNDYNRF 1234	
RESULT 9		
ID	AAW77533 standard; protein; 106 AA.	
XX	AAW77533;	
AC	AAW77533;	
XX	30-OCT-1998 (first entry)	
DT		
XX	Staphylococcus aureus protein of unknown function.	
DE		
XX	Staphylococcus aureus protein; immune response induction; eye infection;	
KM	antibody production; T-cell immune response; gastrointestinal infection;	
KM	respiratory infection; inhibitor; bacterial infection; cardiac infection;	
KW	central nervous system; kidney infection; urinary tract infection;	
KW	antimicrobial compound identification; broad spectrum antibiotic;	
KW	therapy.	
OS	Staphylococcus aureus.	
XX		
PN	EP841394-A2.	
XX	13-MAY-1998.	
PD		
XX	24-SEP-1997; 97EP-00307485.	
PF		
XX	24-SEP-1996; 96US-0027032P.	
PR		
XX	(SMIK ) SMITHKLINE BEECHAM CORP.	
PA	(SMIK ) SMITHKLINE BEECHAM PLC.	
XX		
PI	Black MT, Hodson JE, Knowles DJC, Reichard RW, Nicholas RO;	
PI	Burnham MKR, Prate JM, Rosenberg M, Ward JM, Lonetto MA;	
XX		

```

DR WPI; 1998-252940/23.
DR N-PSDB; AAV53334.
XX New nucleic acid sequences from Staphylococcus aureus WCHU29 - useful in
PT vaccines and for treatment of bacterial infections of e.g. respiratory
PT tract and central nervous system.
XX
PS Claim 11; Page 236; 390pp; English.
XX
CC This sequence represents a Staphylococcus aureus protein of unknown
CC function, and is encoded by a DNA sequence of the invention. The DNA
CC sequences were isolated from Staphylococcus aureus WCHU29 (NCIMB 40771).
CC Host cells containing the DNA sequences are used to produce polypeptides
CC or fragments. The proteins are used in the treatment of disease, for
CC inducing an immune response by administering them, to produce antibody
CC and/or T-cell immune response. Antagonists of the proteins are used for
CC the inhibition of bacterial polypeptides. Conditions which may be treated
CC include bacterial infections, especially respiratory, cardiac,
CC gastrointestinal, central nervous, eye, kidney, urinary tract, skin,
CC bones and joints. The proteins can also be used to identify antimicrobial
CC compounds which are broad spectrum antibiotics, especially useful in the
CC treatment of H. pylori infection
XX
SQ Sequence 106 AA;
XX
Query Match 55.9%; Score 38; DB 2; Length 106;
Best Local Similarity 45.5%; Pred. No. 70;
Matches 5; Conservative 5; Mismatches 1; Indels 0; Gaps 0
QY 3 RRFTESENEY 13
: : : : :
DB 27 KNEYTAKTNEY 37
XX
RESULT 10
ID AAG82630 standard; protein; 133 AA.
XX AAG82630;
XX AAG82630;
XX 03-SEP-2001 (First entry)
XX
DE S. epidermidis open reading frame protein sequence SEQ ID NO:2354.
XX
KW Staphylococcus epidermidis SRI strain; infection; diagnosis; vaccination;
KW endocarditis.
XX
OS Staphylococcus epidermidis.
XX
PN WO200134809-A2.
XX
PD 17-MAY-2001.
XX
PF 09-NOV-2000; 2000WO-US030782.
XX
PR 09-NOV-1999; 99US-0164258P.
XX
PA (GLAXO ) GLAXO GROUP LTD.
XX
PI Kimerly WJ;
XX
DR WPI; 2001-316495/33.
DR N-PSDB; AAV53480.
XX
PT Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
PT useful for vaccinating against infections, e.g. endocarditis.
XX
PS Claim 18; Page 623; 2189pp; English.
XX
AAH53304 to AAH53970 represent nucleic acids (I) encoding polypeptides
CC (II), given in AA681454 to AAG81120, from Staphylococcus epidermidis. (I)
CC and (II) can have antibacterial activity and therefore can be used in
CC vaccination. The nucleic acids (I) may be used to produce the S.

```

epidermidis polypeptides (II) via the production of vectors containing them which are used to produce hosts cells which express the polypeptides. The polypeptides (II) (and/or nucleic acids) may then be used to vaccinate subjects and to raise antibodies against the bacteria. The polypeptides may also be used to assay for other inhibitors of their activity and therefore identify compounds that may be used for the treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to AAH55099 represent specifically claimed S. epidermidis genomic DNA polynucleotide sequences from the present invention. AAH55091 to AAH55098 represent oligonucleotide sequences and primers which are used in the exemplification of the present invention. N.B. The present invention specifically claims all the polynucleotide sequences given in the sequence listing of the present specification, however the sequence listing only goes up to SEQ ID NO:4454 so even though sequences are given in the disclosure for SEQ ID NO:4465 to 4472, no sequences are present for SEQ ID NO:4455 to 4464

Sequence 133 AA;

Query Match 55.9%; Score 38; DB 4; Length 133;  
Best Local Similarity 50.0%; Pred. No. 89;  
Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 LRRKFTSEYNE 12  
:|:|:|:|:|:  
69 MTKSKFNDEYNO 100

RESULT 11

ADCC3861  
ID ADCC3861 standard; protein; 334 AA.

AC ADC3861;

DT 18-DEC-2003 (first entry)

DE Protein sequence (SeqID 128) exhibiting nitrilase activity.

XX enzyme; nitrilase; nitrile; cyanohydrin; ammonia; biocatalyst;

XX enantiomer; chiral medicine.

OS Unidentified.

PN WO200300840-A2.

PD 03-JAN-2003.

PF 15-MAY-2002; 2002WO-US015983.

PR 21-JUN-2001; 2001US-0300189P.

PR 30-JUL-2001; 2001US-0309006P.

PR 22-JAN-2002; 2002US-0351336P.

PA (DIVE-) DIVERSA CORP.

PA (MADD/) MADDEN D.

PI Madden M, Desautis G, Chaplin JA, Weiner DP, Milan A, Chi E;

PI Short JM, Burk M;

XX MPI; 2003-201417/19.

XX N-PSDB; ADC23860.

PT Novel nitrilase polypeptide, useful for making (R) - or (S)-ethyl-4-cyano-3-hydroxybutyric acid or (R) - or (S)-mandelic acid or (S) - or (R)-phenyl lactic acid derivative and for producing pharmaceutical composition, and food additive.

XX Claim 40; SEQ ID NO 128; 560pp; English.  
CC This invention relates to nitrilases and the nucleic acids that encode these enzymes thereof. Specifically, it refers to polypeptides that exhibit nitrilase activity, i.e. the ability to directly hydrolyse nitriles or cyanohydrins into their corresponding carboxylic acids and

ammonia. Nitrilases have commercial utility as biocatalysts for use in the synthesis of enantiomerically pure aromatic and aliphatic amino acids, as well as hydroxy acids, which are important for the development of chiral medicines. Furthermore, the present invention describes nitrilases, isolated from mesophilic microorganisms, that have improved activity and stability at increased pH and temperature. They are also inexpensive, efficient catalysts, have broad substrate specificity and are capable of chiral differentiation. This polypeptide is a protein sequence that exhibits nitrilase activity of the invention.

Sequence 334 AA;

Query Match 55.9%; Score 38; DB 7; Length 334;  
Best Local Similarity 58.3%; Pred. No. 2,3e+02;  
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 LRRKFTSEYNE 12  
:|:|:|:|:|:  
125 LVRRKITPYHE 136

RESULT 12

AAU07861  
ID AAU07861 standard; protein; 685 AA.

AC AAU07861;

DT 18-DEC-2001 (first entry)

DE Polypeptide sequence for mammalian SPg3.

XX Mammalian; reproductive-specific protein; male infertility;

XX spermatogenesis; sperm count disorder; anti infertility; reproduction.

OS Mammalia.

PN WO200166752-A2.

PD 13-SEP-2001.

PF 07-MAR-2001; 2001WO-US007371.

PR 12-MAR-2000; 2000US-0187518P.

PR 12-JAN-2001; 2001US-0261557P.

PA (WHD) WHITEHEAD INST BIOMEDICAL RES.

PA Wang PJ, Page DC;

PA MPI; 2001-570774/64.

PA N-PSDB; AAS13625.

PT Novel reproduction-specific protein, useful for treating disorders of reduced sperm count, enhancing/increasing sperm count and/or sperm activity.

XX Claim 22; Fig 6; 151pp; English.

CC The present invention relates to the isolation of novel mammalian and human reproductive-specific proteins (AAU0785-AAU0789), and the nucleic acids encoding them. The nucleic acids encoding reproductive-specific proteins are useful for diagnosing infertility which is a result of reduced sperm count, reduced sperm motility, malformed sperm or combinations of these. The sequences of the invention are useful as markers for spermatogenic cells, for identifying genes or proteins characteristic of male infertility, diagnosing or aiding in the diagnosis of infertility in men, and for contraception in which sperm production or sperm count is reduced or defective sperm is produced. Antibodies to reproductive-specific proteins are useful for determining the presence of these proteins in a sample obtained from a man being assessed for infertility, for identifying the expression of genes in particular cell type or particular developmental stage, for studies of spermatogenesis, and for immunofluorescence of germ cells or in Western blots for



CC assessing the presence of the protein the antibody binds. The sequences  
 CC of the invention are also useful for treating disorders of reduced sperm  
 CC count, and for increasing sperm count and/or sperm activity. The nucleic  
 CC acids of the invention are useful in gene therapy. AA07859-AA07882  
 CC represent the mammalian reproduction-specific proteins of the present  
 CC invention

SO Sequence 685 AA;

Query Match 55.9%; Score 38; DB 4; Length 685;  
 Best Local Similarity 58.3%; Pred. No. 4.9e+02;  
 Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 2 IRRKFTSEYNEY 13  
 Db 206 VONKFTSEYNEY 217

RESULT 13  
 AA01602  
 ID AA01602 standard; protein; 793 AA.

AC AA01602;  
 DT 16-APR-1997 (first entry)

DE Thermotoga maritima MutS protein.

XX MutS; thermostable protein; bulge loop; heteroduplex; amplification;  
 KM polymerase chain reaction; PCR; ligase chain reaction; LCR; mutation;  
 KM DNA misincorporation; genetic disease; cancer; diagnosis.

XX Thermotoga maritima.

XX WO9639525-A1.

XX 12-DEC-1996.

XX 04-JUN-1996; 96WO-US008677.

XX 06-JUN-1995; 95US-00468558.

XX (UYNY ) UNIV NEW YORK CITY MT SINAI SCHOOL.

XX Wetmur JG;

XX WPI; 1997-043147/04.

XX N-PSDB; AAT58228.

XX Thermostable protein, MutS, and corresponding coding sequences - used in  
 PT amplification methods to detect particular sequences associated with  
 PT genetic diseases and cancer.

XX Claim 13; Fig 4; 93pp; English.

XX The Thermotoga maritima thermostable protein, MutS (AA01602), binds  
 CC specifically to bulge loops in a heteroduplex nucleic acid. Recombinant  
 CC MutS can be produced in transformed host cells using an isolated gene  
 CC sequence (AAT58228). MutS proteins (see also AA01601) are useful in  
 CC methods for reducing DNA misincorporation in amplification reactions such  
 CC as PCR and LCR, in methods for detecting and amplifying nucleic acids  
 CC contg. specific sequences, and in methods for selecting against a nucleic  
 CC acid comprising a specific sequence. These methods enable detection of  
 CC partic. sequences associated with infectious diseases, genetic disorders  
 CC or cellular disorders such as cancer, e.g. oncogenes

XX Sequence 793 AA;

Query Match 55.9%; Score 38; DB 2; Length 793;  
 Best Local Similarity 53.8%; Pred. No. 5.8e+02;  
 Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Qy 1 IIRKFTSEYNEY 13

Db 401 VIKRFSSELDXY 413

RESULT 14

AA03285  
 ID AA03285 standard; protein; 3080 AA.

AC AA03285;

DT 25-MAR-2003 (revised)

DT 17-DEC-2001 (revised)

DT 06-APR-1990 (first entry)

DE Sequence of clone HIV-2 SBL/ISY.

XX HIV-2; proviral clone HIV-2 SBL/ISY.

XX Human immunodeficiency virus 2.

XX USN7331212-N.

XX 29-AUG-1989.

XX 31-MAR-1989; 89US-00331212.

XX 31-MAR-1989; 89US-00331212.

XX (USSH ) US DEPT HEALTH & HUMAN SERVICE.

XX (USDC ) US SEC OF COMMERCE.

XX (USDC ) US SEC OF COMMERCE.

XX Franchini G, Wongschaal F, Gallo R;

XX WPI; 1989-339698/46.

XX N-PSDB; AAN92119.

XX Complete human immuno-deficiency type 2 pro-viral clone - used to

PT generate animal model for function studies of HIV genes in vivo.

XX Disclosure; Fig 5; 43pp; English.

XX The protein is encoded by the third reading frame of HIV-2 SBL/ISY, a

CC proviral clone of HIV-2. (Note: Revised entry submitted to correct the

CC patent number format of US Government-owned NRTS applications to prevent

CC clashes with ongoing US granted patent numbers. For further information

CC please visit the Derwent web site at

CC www.derwent.com/dwpi/updates/nrtis-us.html.) (Updated on 25-MAR-2003 to

CC correct PR field.) (Updated on 25-MAR-2003 to correct PA field.) (Updated

CC on 25-MAR-2003 to correct PI field.)

XX Sequence 3080 AA;

Query Match 55.9%; Score 38; DB 1; Length 3080;  
 Best Local Similarity 45.0%; Pred. No. 2.4e+03;  
 Matches 9; Conservative 2; Mismatches 1; Indels 8; Gaps 1;

Qy 2 IRRKFTSEYNEY 13  
 Db 1986 VRRKRLRLHOTSNEYEW 2005

RESULT 15  
 AA01771  
 ID AA01771 standard; protein; 3210 AA.  
 AC AA01771;  
 DT 25-MAR-2003 (revised)  
 DT 07-NOV-1990 (first entry)  
 XX Deduced sequence encoded by bottom reading frame of cDNA clone HIV-2  
 DB SBL/ISY of HIV related retrovirus strain.

XX HIV vaccine; HIV strain SBL-6669-85.  
 KW  
 XX  
 OS Human immunodeficiency virus.  
 XX  
 PN WO8808449-A.  
 XX  
 PD 03-NOV-1988.  
 XX  
 PF 28-APR-1988; 88MO-SE000218.  
 XX  
 PR 28-APR-1987; 87SE-00001765.  
 XX  
 PA (SBL-) SBL STATENS BAKTERI.  
 PA (STAT-) STATENS BAKTERIOLOGISKA LAB.  
 PA (STAT-) STATENS BAKTERIOLOGISKA LAB.  
 XX  
 PI Albert J, Biberfeld G, Fenyo EM, Norrby E;  
 XX  
 DR WPT; 1988-322769/45.  
 DR N-PSDB; AAN80890.  
 XX  
 PT HIV related human retro-virus strain - used for obtaining antigens for  
 PT assays and vaccines and for prodn. of antibodies for assays.  
 XX  
 PS Claim 9; Fig 4; 28pp; English.  
 XX  
 CC Synthetically produced proteins and peptides, characterised in that the  
 CC AA sequence is derived from the primary nucleotide sequence of HIV-2  
 CC SBL/ISY or a part thereof, or a degenerate thereof are claimed. HIV-2  
 CC SBL/ISY represents the complete genome of the virus SBL-6669 (SBL-6689-  
 CC 85). The proviral DNA was chd. from a genomic library constructed from  
 CC DNA of HUT-78 cells infected with SBL-6669-85 using the lambda-phage  
 CC vector EMBL-3. SBL-6669-85 was isolated from lymphocytes of a West  
 CC African woman. Protection is requested for the entire genome disclosed in  
 CC n80890 and for parts thereof, and corresp. to various genes such as the  
 CC gag gene (corresp. to nucleotides 547 to 2106), the pol gene (nucleotides  
 CC 1827-4931) and the env gene (nucleotides 6144 to 8682), the corresp. AA  
 CC sequences and parts thereof and various products derived therefrom, or  
 CC use thereof, such as clones prepd. by recombinant vector method, HIV test  
 CC devices and methods. X corresponds to the translation of a stop codon.  
 CC (Updated on 25-MAR-2003 to correct PA field.)  
 CC  
 SQ Sequence 3210 AA;

Query Match 55.9%; Score 38; DB 1; Length 3210;  
 Best Local Similarity 45.0%; Pred. No. 2.5e+03;  
 Matches 9; Conservative 2; Mismatches 1; Indels 8; Gaps 1;

OY 2 IRRKF-----TSEYNEY 13  
 :||| :||| :  
 :||| :||| :  
 DB 2030 VRRKLRIRLHQTSEYNEW 2049

Search completed: October 5, 2004, 07:59:48  
 Job time : 14.8904 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2004, 07:36:16 ; Search time 1.94444 Seconds  
(without alignments)  
593.639 Million cell updates/sec

Title: US-09-805-290a-13  
Perfect score: 67  
Sequence: 1 LITRMDKSVNDY 12

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: PIR:\*  
2: PIR:\*  
3: PIR:\*  
4: PIR:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	47	70.1	272	2 T33515	hypothetical prote
2	43	64.2	199	2 E70101	hypothetical prote
3	41	61.2	479	2 D84129	4-hydroxyphenylace
4	41	61.2	562	2 A85042	hypothetical prote
5	40	59.7	532	2 S33758	flavin-containing
6	40	59.7	1659	2 H97926	hypothetical prote
7	40	59.7	1659	2 G95057	endo-beta-N-acetyl
8	39	58.2	191	2 A71714	glucose inhibited
9	39	58.2	590	1 T35257	probable dihydrol
10	39	58.2	689	2 F83843	hypothetical prote
11	39	58.2	851	2 D87252	TonB-dependent rec
12	39	58.2	1191	2 B97116	chromosome segrega
13	39	58.2	172	2 S62761	hypothetical prote
14	38	56.7	192	2 B97270	glucose inhibited
15	38	56.7	343	2 AH1636	N-acetylglutamate
16	38	56.7	1170	2 T44132	plasmid recombinat
17	38	56.7	1179	2 C67922	hypothetical prote
18	38	56.7	1461	2 B70588	probable polyketid
19	37	55.2	175	2 JC7274	DNA binding protei
20	37	55.2	204	2 T21057	hypothetical prote
21	37	55.2	280	2 G89959	conserved hypochet
22	37	55.2	332	2 D83745	3-oxoacyl-lacyl-ca
23	37	55.2	357	2 S21992	envelope protein g
24	37	55.2	414	2 H64203	histidine-tRNA lig
25	37	55.2	460	2 S34947	nitrogenase (EC 1.
26	37	55.2	468	2 T33091	hypothetical prote
27	37	55.2	471	2 T39571	probable regulator
28	37	55.2	529	2 S46116	probable regulator
29	37	55.2	698	2 A84547	hypothetical prote

30	37	55.2	832	2 A31245	neurogenic protein
31	37	55.2	833	2 S19087	gene Delta protein
32	37	55.2	880	2 S08670	neurogenic repeat
33	37	55.2	935	2 S62476	hypothetical prote
34	37	55.2	1278	2 A47462	probable DNA-dirac
35	37	55.2	1781	2 A34374	DNA-directed RNA p
36	36	53.7	196	2 D64829	probable NAD(P)H2
37	36	53.7	210	2 A99487	hypothetical prote
38	36	53.7	244	2 T05140	hypothetical prote
39	36	53.7	262	2 C70587	hypothetical prote
40	36	53.7	326	2 PC4028	dihydroliposamide S
41	36	53.7	327	2 S49796	hypothetical prote
42	36	53.7	341	2 B39654	cell cycle arrest
43	36	53.7	341	2 AD3145	hypothetical prote
44	36	53.7	348	2 T21648	hypothetical prote
45	36	53.7	358	2 S21998	envelope protein g

## ALIGNMENTS

## RESULT 1

T33515

hypothetical protein F59H5.3 - Caenorhabditis elegans

C.Species: Caenorhabditis elegans

C.Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 21-Jul-2000

C.Accession: T33515

R.Jones, K.; Graves, T.; Antoniou, B.

submitted to the EMBL Data Library, October 1998

A.Description: The sequence of C. elegans cosmid F59H5.

A.Reference number: Z21362

A.Accession: T33515

A.Status: preliminary; translated from GB/EMBL/DBJ

A.Molecule type: DNA

A.Residues: 1-272 &lt;CON&gt;

A.Cross-references: EMBL:AF098991; PIR:NAAC67450.1; GSPDB:GN000020; CESP:F59H5.3

A.Experimental source: strain Bristol N2; clone F59H5

C.Genes:

A.Gene: CESP:F59H5.3

A.Map position: 2

A.Intons: 38/2, 168/3, 182/1, 232/3

C.Superfamily: Caenorhabditis elegans

hypothetical protein F59C6.7; POZ domain homology

Query Match 70.1% Score 47; DB 2; Length 272;

Best Local Similarity 77.8%; Pred. No. 1.2;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 4 RMDKSVNDY 12

DB 108 RMDKLVNDY 116

## RESULT 2

E70101

hypothetical protein BB0013 - Lyme disease spirochete

C.Species: Borrelia burgdorferi (Lyme disease spirochete)

C.Date: 13-Feb-1998 #sequence\_revision 13-Feb-1998 #text\_change 08-Oct-1999

C.Accession: E70101

R.Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Jadhava, R.; White,

son, D.; Peterson, J.; Kervatage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt,

Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horat, K.; Roberts, K.; Hatch, B.

Nature 390, 580-586, 1997

A.Authors: Smith, H.O.; Venter, J.C.

A.Title: Genomic sequence of a Lyme disease spirochete, Borrelia burgdorferi.

A.Reference number: A70100; MUID:58065943; PMID:9403685

A.Accession: E70101

A.Status: preliminary; nucleic acid sequence not shown; translation not shown

A.Molecule type: DNA

A.Residues: 1-199 &lt;KLF&gt;

A.Cross-references: GB:AE001115; GB:AE000783; NID:g2687879; PIR:NAAC66401.1; PID:g268789C

A.Experimental source: strain B31

Query Match 64.2% Score 43; DB 2; Length 199;

Best Local Similarity 58.3%; Pred. No. 4.3;  
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;  
QY 1 LITRDKSVNDY 12  
Db 101 IITKCKSINIY 112

## RESULT 3

4-hydroxyphenylacetate-3-hydroxylase BH336 [imported] - *Bacillus halodurans* (strain C-1)  
C/Species: *Bacillus halodurans*  
C/Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 01-Mar-2002  
C/Accession: D84129  
R/Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masu, N.; Fuji, F.; Hira  
Nucleic Acids Res. 28, 4317-4331, 2000  
A/Title: Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and  
A/Reference number: A83650; MUID:20512582; PMID:11058132  
A/Accession: D84129  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-479 <STO>  
A/Cross-references: GB:AP001520; GB:BA000004; NID:g10176401; PIDN:BA07555.1; GSPDB:GN00  
A/Experimental source: strain C-125  
C/Genetics:  
A/Gene: BH336  
C/Superfamily: Escherichia coli 4-hydroxyphenylacetate 3-monooxygenase large chain

Query Match 61.2%; Score 41; DB 2; Length 479;  
Best Local Similarity 50.0%; Pred. No. 25;  
Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;  
QY 1 LITRDKSVNDY 12  
Db 59 LITCWDKXTNEY 70

## RESULT 4

A85042  
hypothetical protein AT4g03310 [imported] - *Arabidopsis thaliana*  
C/Species: *Arabidopsis thaliana* (mouse-ear cress)  
C/Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 17-May-2002  
C/Accession: A85042  
R/Anonymous. The European Union Arabidopsis Genome Sequencing Consortium, The Cold Spring  
Nature 402, 769-777, 1999  
A/Title: Sequence and analysis of chromosome 4 of the plant *Arabidopsis thaliana*.  
A/Reference number: A85001; MUID:20083488; PMID:10617198  
A/Accession: A85042  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-562 <STO>  
A/Cross-references: GB:NC\_001268; NID:g7270201; PIDN:CA877816.1; GSPDB:GN00140  
A/Genetics:  
A/Gene: AT4g03310  
A/Map position: 4  
C/Superfamily: Arabidopsis hypothetical protein FN22.18

Query Match 61.2%; Score 41; DB 2; Length 562;  
Best Local Similarity 65.7%; Pred. No. 30;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 5 WDKSVND 11  
Db 149 WDKSIND 155

## RESULT 5

S33758  
flavin-containing monooxygenase 1 - rat  
C/Species: *Rattus norvegicus* (Norway rat)  
C/Date: 06-Jan-1995 #sequence\_revision 06-Jan-1995 #text\_change 11-Jun-1999  
C/Accession: S33758  
R/Itoh, K.; Kimura, T.; Yokoi, T.; Itoh, S.; Kamataki, T.

Biochim. Biophys. Acta 1173, 165-171, 1993  
A/Title: Rat liver flavin-containing monooxygenase (FMO): cDNA cloning and expression in  
A/Reference number: S33758; MUID:93277949; PMID:8504165  
A/Accession: S33758  
A/Status: preliminary  
A/Molecule type: mRNA  
A/Residues: 1-532 <IT0>  
A/Cross-references: EMBL:M84719; NID:g204151; PIDN:AAA41165.1; PID:g204152  
C/Superfamily: dimethylanthiline monooxygenase (N-oxide-forming)

Query Match 59.7%; Score 40; DB 2; Length 532;  
Best Local Similarity 50.0%; Pred. No. 42;  
Matches 5; Conservative 5; Mismatches 0; Indels 0; Gaps 0;  
QY 1 LITRDKSVNDY 10  
Db 488 IITQWDRVYN 497

## RESULT 6

H97926  
hypothetical protein SP0440 [imported] - *Streptococcus pneumoniae* (strain R6)  
C/Species: *Streptococcus pneumoniae*  
C/Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 22-Oct-2001  
C/Accession: H97926  
R/Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.; Be  
e, R.; Lepland, D.J.; Lee, L.N.; Letkowitz, E.J.; Lu, J.; Matsushima, P.; McAnen, S.; Mc  
Y, P.; Sun, P.M.; Winkler, M.E.  
J. Bacteriol. 183, 5709-5717, 2001  
A/Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;  
A/Title: Genome of the Bacterium *Streptococcus pneumoniae* strain R6.  
A/Reference number: A97872; MUID:21429245; PMID:11544234  
A/Accession: H97926  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-1659 <XUR>  
A/Cross-references: GB:AE007317; PIDN:AAK9244.1; PID:g15458007; GSPDB:GN00174  
C/Genetics:  
A/Gene: SP0440

Query Match 59.7%; Score 40; DB 2; Length 1659;  
Best Local Similarity 75.0%; Pred. No. 1.5e+02;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
QY 5 WDKSVNDY 12  
Db 414 WDKRVNDY 421

## RESULT 7

G95057  
endo-beta-N-acetylglucosaminidase, probable [imported] - *Streptococcus pneumoniae* (strain  
C/Species: *Streptococcus pneumoniae*  
C/Date: 03-Aug-2001 #sequence\_revision 03-Aug-2001 #text\_change 03-Aug-2001  
C/Accession: G95057  
R/Tetzelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heide  
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtapple, E  
nson, T.; Hickey, E.K.; Holt, I.E.  
Science 293, 498-506, 2001  
A/Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,  
A/Title: Complete Genome Sequence of a virulent isolate of *Streptococcus pneumoniae*.  
A/Reference number: A95000; MUID:21357209; PMID:11463916  
A/Accession: G95057  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-1659 <XUR>  
A/Cross-references: GB:AE005672; PIDN:AAK74656.1; PID:g14971970; GSPDB:GN00164; TIGR:SP46  
A/Experimental source: strain TIGR4  
C/Genetics:  
A/Gene: SP0498

Query Match 59.7%; Score 40; DB 2; Length 1659;  
Best Local Similarity 75.0%; Pred. No. 1.5e+02;

Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
QY 5 WDKSVNDY 12  
|||:|  
Db 414 WDKXNDY 421

## RESULT 8

A:71714  
glucose inhibited division protein B (gidB) RP057 - Rickettsia prowazekii  
C:Species: Rickettsia prowazekii  
C:Date: 21-Nov-1998 #sequence\_revision 21-Nov-1998 #text\_change 03-Nov-2000  
C:Accession: A71714  
R:Andersson, S.G.E.; Zomorodipour, A.; Andersson, J.O.; Sichteritz-Ponten, T.; Alsmark, U.  
Nature 396, 133-140, 1998  
A:Title: The genome sequence of Rickettsia prowazekii and the origin of mitochondria.  
A:Reference number: A71630; MUID:99039499; PMID:9823893  
A:Accession: A71714  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-191 <AND>  
A:Cross-references: GB:AJ235270; GB:AJ235269; NID:g3860572; PIDN:CA14528.1; PID:g386062  
A:Experimental source: strain Madrid E  
C:Genetics:  
A:Gene: gidB; RP057  
C:Superfamily: gidB protein

Query Match 58.2%; Score 39; DB 2; Length 191;  
Best Local Similarity 60.0%; Pred. No. 21;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 LITRWDKSVND 10  
|||:|  
Db 17 LITKWNKSN 26

## RESULT 9

T35297  
probable dihydrolipoamide S-succinyltransferase (BC 2.3.1.61) SCSF7.20 [similarity] - St  
C:Species: Streptomyces coelicolor  
C:Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 08-Sep-2000  
C:Accession: T35297  
R:Seeger, K.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Raftery, M.A.  
submitted to the EMBL Data Library, July 1999  
A:Reference number: Z21574  
A:Accession: T35297  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-590 <SEE>  
A:Cross-references: EMBL:AU096872; PIDN:CAB51265.1; GSPDB:GN00070; SCOEDB:SCSF7.20  
A:Experimental source: strain A3(2)  
C:Genetics:  
A:Gene: sucB; SCOEDB:SCSF7.20  
C:Superfamily: Mycobacterium probable dihydrolipoamide succinyltransferase; lipoyl/biotin  
C:Keywords: acyltransferase; coenzyme A  
C:Keywords: lipoyl/biotin-binding homology <LBP1>  
F:4-77/Domain: lipoyl/biotin-binding homology <LBP2>  
F:11-204/Domain: lipoyl/biotin-binding homology <LBP3>  
F:43/Binding site: lipamide (lys) (covalent) #status predicted  
F:170/Binding site: lipamide (lys) (covalent) #status predicted

Query Match 58.2%; Score 39; DB 1; Length 590;  
Best Local Similarity 70.0%; Pred. No. 71;  
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 ITRWDKSVND 11  
|||:|  
Db 146 VTRWMLKSVND 155

## RESULT 10

F83843  
hypothetical protein BH1550 [imported] - Bacillus halodurans (strain C-125)

C:Species: Bacillus halodurans  
C:Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 15-Jun-2001  
C:Accession: F83843  
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fuji, F.; Hirai  
Nucleic Acids Res. 28, 4317-4331, 2000  
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and  
A:Reference number: A83650; MUID:20512582; PMID:11058132  
A:Accession: F83843  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-689 <STO>  
A:Cross-references: GB:AP001512; GB:BA000004; NID:g10174030; PIDN:BA805269.1; GSPDB:GN001  
A:Experimental source: strain C-125  
C:Genetics:  
A:Gene: BH1550

Query Match 58.2%; Score 39; DB 2; Length 689;  
Best Local Similarity 54.5%; Pred. No. 84;  
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 LITRWDKSVND 11  
|||:|  
Db 304 LTRWMDPSVND 314

## RESULT 11

D87252  
tonB-dependent receptor [imported] - Caulobacter crescentus  
C:Species: Caulobacter crescentus  
C:Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 14-Sep-2001  
C:Accession: D87252  
R:Nierman, W.C.; Peltdlym, T.V.; Paulsen, I.T.; Nelson, K.R.; Eisen, J.; Heidelberg, J.  
B.; Laub, M.T.; Deboy, R.T.; Dodson, R.U.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon  
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.  
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001  
A:Title: Complete genome sequence of Caulobacter crescentus.  
A:Reference number: A87249; MUID:21173698; PMID:11259647  
A:Accession: D87252  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-851 <STO>  
A:Cross-references: GB:AE005673; NID:g1342118; PIDN:AAK22016.1; GSPDB:GN00148  
C:Genetics:  
A:Gene: CC0028  
C:Superfamily: ferrichrome-iron receptor 1; tonB-dependent receptor amino-terminal homolo

Query Match 58.2%; Score 39; DB 2; Length 851;  
Best Local Similarity 54.5%; Pred. No. 11e+02;  
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 2 ITRWDKSVNDY 12  
|||:|  
Db 407 VTRKSLNDY 417

## RESULT 12

B97116  
chromosome segregation SMC protein, Arpase [imported] - Clostridium acetobutylicum  
C:Species: Clostridium acetobutylicum  
C:Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 14-Sep-2001  
C:Accession: B97116  
R:Nolling, J.; Breton, G.; Onelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,  
J.; Bacteriol. 183, 4823-4838, 2001  
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo  
A:Reference number: A96900; MUID:21359325; PMID:21359325  
A:Accession: B97116  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1191 <KUN>  
A:Cross-references: GB:AE001437; PIDN:AAK79717.1; PID:g15024720; GSPDB:GN00168  
A:Experimental source: Clostridium acetobutylicum ATCC824  
C:Genetics:

A:Gene: CAC1751

Query Match 58.2%; Score 39; DB 2; Length 1191;  
Best Local Similarity 50.0%; Pred. No. 1.5e+02;  
Matches 6; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 LITRMDKSVNDY 12  
|||:|:  
Db 373 LIEWEKSIKQY 384

#### RESULT 13

662761  
Hypothetical protein 172 - red alga (Cyanidium caldarium) mitochondrion

C:Species: mitochondrion Cyanidium caldarium  
C:Date: 19-Mar-1997 #sequence\_revision 09-May-1997 #text\_change 05-Nov-1999  
C:Accession: S62761

R:Viehmann, S.  
submitted to the EMBL Data Library, March 1995

A:Reference number: S62757  
A:Accession: S62761  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-172 <VIE>  
A:Cross-references: EMBL:Z48930; NID:G791089; PIDN:CA88771.1; PID:G758177  
C:Genetics:  
A:Genome: mitochondrion  
C:Keywords: mitochondrion

Query Match 56.7%; Score 38; DB 2; Length 172;  
Best Local Similarity 70.0%; Pred. No. 27;  
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 LITRMDKSVN 10  
:|:|:|:  
Db 1 MITRWLKKVN 10

#### RESULT 14

E97710

glucose inhibited division protein B [imported] - Rickettsia conorii (strain Malish 7)

C:Species: Rickettsia conorii  
C:Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 22-Oct-2001  
C:Accession: E97710

R:Ogata, H.; Audic, S.; Renesto-Audiffren, P.; Fournier, P.E.; Barbe, V.; Samson, D.; Ro  
science 293, 2093-2098, 2001

A:Title: Mechanisms of Evolution in Rickettsia conorii and Rickettsia prowazekii.  
A:Reference number: A97700; MUID:21442074; PMID:11537893

A:Accession: E97710  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-192 <KUR>  
A:Cross-references: GB:AE006914; PIDN:AAL02623.1; PID:G15619123; GSPDB:GN00173  
C:Genetics:  
A:Gene: gidB  
C:Superfamily: gidB protein

Query Match 56.7%; Score 38; DB 2; Length 192;  
Best Local Similarity 50.0%; Pred. No. 31;  
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 LITRMDKSVN 10  
:|:|:|:  
Db 18 LVKKMNKSN 27

#### RESULT 15

AH1636

N-acetylglutamate gamma-semialdehyde dehydrogenases homolog argC [imported] - Listeria

C:Species: Listeria innocua  
C:Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 14-Dec-2001  
C:Accession: AH1636

R:Glaser, P.; Frangoul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker

.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.  
D.; Jones, L.M.; Karst, U.  
Science 294, 849-853, 2001  
A:Authors: Kref, U.; Kuhn, M.; Kunet, F.; Kurapat, G.; Madueno, E.; Maltournam, A.; Ma  
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, U.A.; Voss, H.; Wehland,  
A:Title: Comparative genomics of Listeria species.  
A:Reference number: AB1077; MUID:21537279; PMID:11679669  
A:Accession: AH1636

A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-343 <GLA>  
A:Cross-references: GB:AL592022; PIDN:CAC96864.1; PID:G16414120; GSPDB:GN00178  
A:Experimental source: strain Clp11262  
C:Genetics:  
A:Gene: argC  
C:Superfamily: N-acetyl-gamma-glutamyl-phosphate reductase

Query Match 56.7%; Score 38; DB 2; Length 343;  
Best Local Similarity 62.5%; Pred. No. 58;  
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 2 ITRMDKSV 9  
:|:|:|:  
Db 216 LTKMDKSI 223

Search completed: October 5, 2004, 08:17:02  
Job time: 5.94444 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2004, 07:29:35 ; Search time 1.07407 Seconds  
(without alignments)  
581.749 Million cell updates/sec

Title: US-09-805-290A-13

Perfect score: 67

Sequence: 1 LITRMDKSVNDY 12

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	43	64.2	199	Y013 BORBU	O51046 borrelia bu
2	40	59.7	532	FM01 RAT	P36365 rattus norv
3	39	58.2	191	GIDB_RICPR	Q92889 rickettsia
4	39	58.2	1501	NINC_DROME	P10567 drosophila
5	38	56.7	192	GIDB_RICCN	Q92422 rickettsia
6	38	56.7	248	YH62_FUSIN	Q8157 fusobacteri
7	38	56.7	343	ARGC_LISIN	Q925b7 listeria in
8	38	56.7	420	PRE2_STIAM	P22490 staphylococ
9	38	56.7	1179	VG56_HSV1	Q00099 ictaluriid h
10	37	55.2	332	FAH2_BACBD	O91875 bacillus ba
11	37	55.2	414	SIH_FICGS	P47281 mycoplasma
12	37	55.2	460	ANFK_RHOCA	Q07935 rhodobacter
13	37	55.2	529	YB89_YEAST	P38140 saccharomyc
14	37	55.2	749	PEX_HUMAN	P78562 homo sapien
15	37	55.2	749	PEX_MOUSE	P70669 mus muscula
16	37	55.2	833	DL_DROME	P10041 drosophila
17	37	55.2	925	RNT1_SCHPO	O09850 schizosach
18	37	55.2	1744	REAL_TRYIB	P16355 trypanosoma
19	36	53.7	196	YCAK_ECOLI	P43340 escherichia
20	36	53.7	262	YN72_MYCTU	O05826 mycobacteri
21	36	53.7	327	SDS3_YEAST	P05055 saccharomyc
22	36	53.7	341	BUB3_YEAST	P64449 saccharomyc
23	36	53.7	389	ACTUC_STIAM	O91c9 staphylococ
24	36	53.7	389	ACUC_STIAM	Gnm34 staphylococ
25	36	53.7	398	PSY_DAUCA	Q3588 daucus caro
26	36	53.7	460	ENGA_CAVOE	Q92186 campylobact
27	36	53.7	610	YEM6_YEAST	P40024 saccharomyc
28	36	53.7	667	EM70_YEAST	P32802 saccharomyc
29	36	53.7	766	GAP1_SCHPO	P33277 echizosach
30	36	53.7	799	Y231_BUCAP	O81965 buchiera ap
31	36	53.7	851	ENV_HV1B8	P04582 human immun
32	36	53.7	851	ENV_HV1M7	P19551 human immun
33	36	53.7	856	ENV_HV1B1	P03375 human immun

34	36	53.7	856	1	ENV_HV1H2	P04578 human immun
35	36	53.7	856	1	ENV_HV1H3	P04624 human immun
36	36	53.7	856	1	ENV_HV1M	Q70626 human immun
37	36	53.7	856	1	ENV_HV1PV	P03376 human immun
38	36	53.7	856	1	ENV_HV1BR	P03377 human immun
39	36	53.7	856	1	ENV_HV1BP	P03378 human immun
40	36	53.7	856	1	ENV_HV1BP	P03379 human immun
41	36	53.7	856	1	ENV_HV1BP	P03380 human immun
42	36	53.7	856	1	ENV_HV1BP	P03381 human immun
43	36	53.7	856	1	ENV_HV1BP	P03382 human immun
44	36	53.7	856	1	ENV_HV1BP	P03383 human immun
45	36	53.7	856	1	ENV_HV1BP	P03384 human immun

#### ALIGNMENTS

RESULT 1  
Y013 BORBU STANDARD; PRT; 199 AA.  
ID Y013 BORBU  
AC O51046;  
DT 15-DEC-1998 (Rel. 37, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hypothetical protein BB0013.  
GN BB0013.  
OS Borrelia burgdorferi (Lyme disease spirochete).  
OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia.  
OX NCBI\_TaxID=139;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 35210 / B31;  
RX MEDLINE=96065943; PubMed=9403685;  
RA Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,  
RA Lathigra R., White O., Ketchum K.A., Dodson R., Hickey E.K., Gwinn M.,  
RA Dougherty B., Tomb J.-F., Fleischmann R.D., Richardson D.,  
RA Peterson J., Kierulff A.R., Quackenbush J., Salzberg S., Hanson M.,  
RA van Vugt R., Palmer N., Adams M.D., Gocayne J.D., Weidman C.,  
RA Uitterlind T., White O., Cotton M.D., Horst K., Roberts K., Hatch B.,  
RA Gairard S., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B.,  
RA Smith H.O., Venter J.C.;  
RA "Genomic sequence of a Lyme disease spirochete, Borrelia  
burgdorferi.";  
RL Nature 390:580-586 (1997).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb.ch/announce/>  
CC or send an email to [license@isb.ch](mailto:license@isb.ch)).  
CC -----  
CC EMBL; AE001115; AAC66401.1; --  
DR PIR; E70101; E70101.  
DR TIGR; BB0013; --  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 199 AA; 23301 MW; C645D1603BDEP610 CRC64;  
Query Match 64.2%; Score 43; DB 1; Length 199;  
Best Local Similarity 58.3%; Pred. No. 2.4;  
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;  
OY 1 LITRMDKSVNDY 12  
Db 101 LITRMDKSVNDY 112  
RESULT 2  
FM01 RAT STANDARD; PRT; 532 AA.  
ID FM01 RAT  
AC P36365;  
DT 01-JUN-1994 (Rel. 29, Created)

```

DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Dimethylamine monooxygenase [N-oxide forming] 1 (EC 1.14.13.8)
DE (Hepatic flavin-containing monooxygenase 1) (FMO 1) (Dimethylamine
DE oxidase 1).
GN FMO1 OR FMO-1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Liver;
RX MEDLINE=93277949; PubMed=8504165;
RA Itoh K., Kimura T., Yokoi T., Itoh S., Kametaki T.;
RT "Rat liver flavin-containing monooxygenase (FMO): cDNA cloning and
RT expression in yeast.";
RL Biochim. Biophys. Acta 1173:165-171(1993).
RN [2]
RP SEQUENCE OF 34-42 AND 52-56.
RC STRAIN=Sprague-Dawley; TISSUE=Liver;
RX MEDLINE=97248495; PubMed=9094723;
RA Hay J.C., Chao D.S., Kuo C.S., Scheller R.H.;
RT "Protein interactions regulating vesicle transport between the
RT endoplasmic reticulum and Golgi apparatus in mammalian cells.";
RL Cell 89:149-158(1997).
CC -1- FUNCTION: This protein is involved in the oxidative metabolism of
CC a variety of xenobiotics such as drugs and pesticides. Form I
CC catalyzes the N-oxygenation of secondary and tertiary amines.
CC -1- CATALYTIC ACTIVITY: N,N-dimethylaniline + NADPH + O(2) = N,N-
CC dimethylaniline N-oxide + NADP(+) + H(2)O.
CC -1- COFACTOR: FAD.
CC -1- SUBCELLULAR LOCATION: Microsomal.
CC -1- TISSUE SPECIFICITY: Liver.
CC -1- SIMILARITY: Belongs to the FMO family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M84719; AAA41165.1; -.
DR PIR: S33758; S33758.
DR InterPro: IPR000960; Flav_cont_mnoxgn.
DR Pfam: PF00743; FMO-like, 1.
DR PRINTS: PR00370; FMOXYGENASE.
DR Oxidoreductase; Monooxygenase; NADP; Flavoprotein; FAD; Microsome;
KW Transmembrane; Multigene family.
FT NP-BIND 9
FT NP-BIND 14 NADP (ADP PART) (POTENTIAL).
FT NP-BIND 191 196 NADP (POTENTIAL).
SQ SEQUENCE 532 AA; 59781 MW; A8195C1E973DF92 CRC64;
Query Match 59.7%; Score 40; DB 1; Length 532;
Best Local Similarity 50.0%; Pred. No. 22;
Matches 5; Conservative 5; Mismatches 0; Indels 0; Gaps 0;
QY 1 LITRWDSDVN 10
DB 488 LITRWDSDVN 497

```

```

GN G1DB OR RP057.
OS Rickettsia prowazekii.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Rickettsiaceae; Rickettsieae; Rickettsia.
OX NCBI_TaxID=782;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Madrid B;
RX MEDLINE=99039495; PubMed=9823893;
RA Andersson S.G.B., Zomrodipour A., Andersson J.O.,
RA Scharitz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,
RA Eriksson A.-S., Winkler H.H., Kurland C.G.;
RT "The genome sequence of Rickettsia prowazekii and the origin of
RT mitochondria.";
RL Nature 396:133-140(1998).
CC -1- FUNCTION: Probable S-adenosyl-L-methionine dependent
CC methyltransferase specific for a sterol and/or lipid substrate (By
CC similarity).
CC -1- SIMILARITY: Belongs to the g1db family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AJ235270; CA414528.1; -.
DR PIR: A71714; A71714.
DR HAMAP: MF_00074; -.
DR InterPro: IPR003682; G1DB.
DR Pfam: PF02527; G1DB; 1.
DR ProDom: PD004441; G1DB; 1.
DR TIGRFAMs: TIGR00138; G1DB; 1.
KW Transferase; Methyltransferase; Complete proteome.
SQ SEQUENCE 191 AA; 21830 MW; 7A15026A16CDD6C1 CRC64;
Query Match 58.2%; Score 39; DB 1; Length 191;
Best Local Similarity 60.0%; Pred. No. 11;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
QY 1 LITRWDSDVN 10
DB 17 LITRWDSDVN 26

```

```

RESULT 4
NINQ DROME STANDARD; PRT; 1501 AA.
AC P10676; P10677;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Neither inactivation nor afterpotential protein C (EC 2.7.1.37).
GN NINAC.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A., FUNCTION, AND ALTERNATIVE SPLICING.
RX MEDLINE=88151067; PubMed=2449973;
RA Montell C., Rubin G.M.;
RT "The Drosophila ninac locus encodes two photoreceptor cell specific
RT proteins with domains homologous to protein kinases and the myosin
RT heavy chain head.";
RL Cell 52:757-772(1988).
CC -1- FUNCTION: Required for photoreceptor cell function. The ninac
CC proteins combines putative serine/threonine-protein kinase and
CC myosin activities.
CC -1- CATALYTIC ACTIVITY: ATP + a protein = ADP + a phosphoprotein.

```



CC -1- SUBCELLULAR LOCATION: Cytoskeleton.  
 CC -1- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=2;  
 CC Name=Long;  
 CC IsoId=P10676-1; Sequence=Displayed;  
 CC Name=Short;  
 CC IsoId=P10676-2; Sequence=VSP\_004940, VSP\_004941;  
 CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE SER/THR  
 CC FAMILY OF PROTEIN KINASES.  
 CC -1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE MYOSIN  
 CC SUPERFAMILY.  
 CC -1- SIMILARITY: Contains 1 IQ domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; J03131; AAA28718.1; -  
 CC EMBL; J03131; AAA28719.1; -  
 CC EMBL; M20230; AAA28721.1; -  
 CC EMBL; M20231; AAA28720.1; -  
 CC PIR; A29813; A29813.  
 CC PIR; B29813; B29813.  
 CC HSSP; P08799; 1MND.  
 CC FLYBASE; FBgn0002338; ninaC.  
 CC DR GO; GO:0005737; C:cytoplasm; IDA.  
 CC DR GO; GO:0042385; C:myosin III; NAS.  
 CC DR GO; GO:0016028; C:rhodomere; IDA.  
 CC DR GO; GO:0005516; F:calmodulin binding; IMP.  
 CC DR GO; GO:0004674; F:protein serine/threonine kinase activity; IDA.  
 CC DR GO; GO:0016062; P:adaptation of rhodopsin mediated signaling; IMP.  
 CC DR GO; GO:0007010; P:cytoskeleton organization and biogenesis; IMP.  
 CC DR GO; GO:0016059; P:deactivation of rhodopsin mediated signaling; IMP.  
 CC DR GO; GO:0007603; P:phototransduction, visible light; IMP.  
 CC DR GO; GO:0008104; P:protein localization; IMP.  
 CC DR InterPro; IPR000048; IQ\_region.  
 CC DR InterPro; IPR001609; myosin\_head.  
 CC DR InterPro; IPR000719; Prot\_kinase.  
 CC DR InterPro; IPR002290; Ser\_thr\_kinase.  
 CC DR InterPro; IPR008266; Tyr\_kinase\_AS.  
 CC DR Pfam; PF00612; IQ\_2.  
 CC DR Pfam; PF00063; myosin\_head; 1.  
 CC DR Pfam; PF00069; pkinase; 1.  
 CC DR PRINTS; PR00193; MYOSINHEAVY.  
 CC DR ProDom; PD000355; myosin\_head; 1.  
 CC DR ProDom; PD000001; Prot\_kinase; 1.  
 CC DR SMART; SM00015; IQ\_2.  
 CC DR SMART; SM000242; MYSC; 1.  
 CC DR SMART; SM00220; S\_TKC; 1.  
 CC DR PROSITE; PS50096; IQ\_1.  
 CC DR PROSITE; PS00107; PROTEIN\_KINASE\_ACP; FALSE\_NEG.  
 CC DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; 1.  
 CC DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
 CC DR CytoSkeleton; Actin-binding; Myosin; Transferase; Vision;  
 CC Serine/threonine-protein kinase; Alternative splicing.  
 CC -----  
 CC DR DOMAIN 16 282  
 CC FT DOMAIN 335 1035  
 CC FT DOMAIN 1036 1065  
 CC FT DOMAIN 1066 1501  
 CC FT NP\_BIND 22 30  
 CC FT BINDING 45 45  
 CC FT ACT\_SITE 145 145  
 CC FT DOMAIN 913 934  
 CC FT VAAPSLIC 1082 1135  
 CC -----  
 CC VAAPSLIC 1136 1501  
 CC Missing (in isoform Short).  
 CC -----

FT CONFLICT 253 253 /FTId=VSP\_004941.  
 FT CONFLICT 1089 1089 K -> Q (IN REF. 1; AAA28720/AAA28721).  
 FT CONFLICT 1501 AA; 174269 MW; D167EABCG2A3933A CRC64;  
 SQ SEQUENCE  
 Query Match 58.2%; Score 39; DB 1; Length 1501;  
 Best Local Similarity 50.0%; Pred. No. 95;  
 Matches 5; Conservative 2; Mismatches 3; Indels 0; Gaps 0;  
 QY 3 TRWDKSYNDY 12  
 DB 249 TWMSKQINDP 258  
 RESULT 5  
 ID GIBB\_RICCN STANDARD; PRT; 192 AA.  
 AC Q92U12;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Methyltransferase gldb (EC 2.1.1.-) (Glucose inhibited division  
 DE protein B).  
 GN GIBB OR RC0085.  
 OS Rickettsia conorii.  
 CC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;  
 CC Rickettsiaceae; Rickettsiae; Rickettsia.  
 CC NCB1\_taxid=781;  
 CC (1)  
 CC SEQUENCE FROM N.A.  
 CC STRAIN=Malish 7;  
 CC MEDLINE=21442074; PubMed=11557893;  
 CC Ogata H., Audic S., Renesto-Audiffren P., Fournier P.-E., Barbe V.,  
 CC Samson D., Roux V., Cossart P., Weissenbach J., Claverie J.-M.,  
 CC Raoult D., Mechanisms of evolution in Rickettsia conorii and R. prowazekii.",  
 CC RT Science 293:2093-2098(2001).  
 CC RL -1- FUNCTION: Probable S-adenosyl-L-methionine dependent  
 CC methyltransferase specific for a sterol and/or lipid substrate (By  
 CC similarity).  
 CC -1- SIMILARITY: Belongs to the gldb family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; AB008575; AAL02623.1; -  
 CC DR PIR; E97710; E97710.  
 CC DR HAMAP; MF\_00074; -; 1.  
 CC DR InterPro; IPR003682; Gldb.  
 CC DR Pfam; PF02527; Gldb; 1.  
 CC DR ProDom; PD004411; Gldb; 1.  
 CC DR TIGRFAMs; TIGR00138; gldb; 1.  
 CC DR Transferase; Methyltransferase; Complete proteome.  
 CC KW SEQUENCE 192 AA; 22084 MW; ED1285DE596CEBC CRC64;  
 SQ  
 Query Match 56.7%; Score 38; DB 1; Length 192;  
 Best Local Similarity 50.0%; Pred. No. 17;  
 Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 LTRWDKSYN 10  
 DB 18 LVKMKNSIN 27  
 RESULT 6  
 YH62\_FUSNN STANDARD; PRT; 248 AA.  
 ID YH62\_FUSNN  
 AC Q8RI57;

```

DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical UPF0246 protein FN1762.
GN FN1762.
OS Fusobacterium nucleatum (subsp. nucleatum).
OC Bacteria; Fusobacteria; Fusobacteriales; Fusobacteriaceae;
OC Fusobacterium.
OX NCBI_TaxID=76856;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 25586; PubMed=11889109;
RX MEDLINE=21886394;
RA Kapatal V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A.,
RA Bhattacharya A., Barham A., Gardner W., Grechkin G., Zhu L.,
RA Varsava O., Chu L., Kogan Y., Chaga O., Goldsman E., Bernal A.,
RA Larsen N., P'Souza M., Walunas T., Pusch G., Haselkorn R.,
RA Fournier N., Kyprides N., Overbeek R.;
RT "Genome sequence and analysis of the oral bacterium Fusobacterium
RT nucleatum strain ATCC 25586."
RL J. Bacteriol. 184:2005-2018 (2002).
CC -1- SIMILARITY: Belongs to the UPF0246 family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AB010480; AAL93875.1; -
CC DR HAMAP; MF_00652; -; 1.
CC DR InterPro; IPR005583; DUF328.
CC DR Pfam; PF03883; DUF328; 1.
CC KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 248 AA; 29228 MW; 8A0D599131C5CDBF CRC64;

Query Match
Best Local Similarity 56.7%; Score 38; DB 1; Length 248;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 5 WDKSVNDY 12
DB 142 WKQVNDY 149

RESULT 7
ID ARGC LISIN STANDARD; PRT; 343 AA.
AC Q92BB7;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE N-acetyl-gamma-glutamyl-phosphate reductase (EC 1.2.1.38) (AGPR) (N-
DE acetyl-glutamate semialdehyde dehydrogenase) (NAGSA dehydrogenase).
GN ARGC OR L1N163.
OS Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
OX NCBI_TaxID=1642;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CIP 11262; Serovar 6a;
RX MEDLINE=21537279; PubMed=11679669;
RA Glaser P., Fangel L., Buchrieser C., Ruenick C., Amend A.,
RA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,
RA Charbit A., Checroun F., Couve E., de Darvar A., Dehoux P.,
RA Domian E., Dominguez-Bernal G., Duchaud E., Durant L., Dussauget O.,
RA Entian K.-D., Fath H., Garcia-del Portillo F., Garrido P.,
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA Jones L.-M., Kaerst U., Kreft J., Kuhn M., Kunst F., Kurapkia D.,
RA Madueno E., Matounian A., Mata Vicente U., Ng E., Negari H.,
RA Nordstiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purrell R.,

```

```

RA Remmel B., Rose M., Schluter T., Simoes N., Tierrez A.,
RA Vazquez-Boland J.-A., Voss H., Wehlend U., Cossart P.;
RT "Comparative genomics of Listeria species."
RL Science 294:849-852 (2001).
CC -1- CATALYTIC ACTIVITY: N-acetyl-L-glutamate 5-semialdehyde + NADP(+)
CC + phosphate = N-acetyl-5-glutamyl phosphate + NADPH.
CC -1- PATHWAY: Arginine biosynthesis; third step.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -1- SIMILARITY: Belongs to the NAGSA dehydrogenase family. Subfamily
CC 1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AL596169; CAC96864.1; -
CC DR PIR; AH1636; AH1636.
CC DR Listlist; L1N01633; -
CC DR HAMAP; MF_00150; -; 1.
CC DR InterPro; IPR007076; AGPR_act_site.
CC DR InterPro; IPR000534; Semialdh_ch.
CC DR Pfam; PF01118; Semialdehyde_dh; 1.
CC DR Pfam; PF02774; Semialdehyde_dhc; 1.
CC DR Prodom; PD003765; AGPR_act_site; 1.
CC DR PROSITE; PS01224; ARGC_1;
CC KW Arginine biosynthesis; Oxidoreductase; NADP; Complete proteome.
CC FT SITE 147 147
CC ACCT SITE 147 147
CC FT SITE 147 147
CC KW Arginine biosynthesis; Oxidoreductase; NADP; Complete proteome.
SQ SEQUENCE 343 AA; 37643 MW; 05E473821EAA6AE7 CRC64;

Query Match
Best Local Similarity 62.5%; Score 38; DB 1; Length 343;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 2 ITRWDKSV 9
DB 216 ITRWDKSI 223

RESULT 8
ID PRE2 STAMM STANDARD; PRT; 420 AA.
AC P22490; Q52091;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Plasmid recombination enzyme (Mobilization protein).
GN PRE OR MOB OR S4V0031 OR S40029.
OS Staphylococcus aureus (strain Mu50 / ATCC 706699),
OS Staphylococcus aureus (strain N315), and
OS Staphylococcus aureus.
OG Plasmid PUB110.
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=158878, 158879, 1280;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Mu50 / ATCC 706699, and N315;
RX MEDLINE=21311952; PubMed=1148146;
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Iian J.-O., Ito T.,
RA Kanamori M., Matsumaru H., Matuyama A., Murakami H., Hoshiyama A.,
RA Mizutani U. Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,
RA Sekimizu K., Hiraoka H., Kihara S., Goto S., Yabuzaki J.,
RA Katsura M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,
RA Hattoni M., Ogasawara N., Hayashi H., Hiratsugu K.;
RT "Whole genome sequencing of methicillin-resistant Staphylococcus
RT aureus."
RL Lancet 357:1225-1240 (2001).
RN [2]
RP SEQUENCE FROM N.A.

```

```

RC PLASMID=pub110;
RX MEDLINE=86301853; PubMed=3744038;
RA Bashkirtov V.I., M.I., Shina N.V., Prozorov A.A.;
RT "Nucleotide sequence and physical map of kanamycin-resistant plasmid
RL Pub110 from Staphylococcus aureus.";
RN Genetika 22:1081-1092(1986).
RP [3]
RP SEQUENCE FROM N.A.
RC PLASMID=pub110;
RX MEDLINE=86206447; PubMed=3010356;
RA McKenzie T., Hoshino T., Tanaka T., Sueoka N.;
RT "The nucleotide sequence of pub110: some salient features in relation
RL to replication and its regulation.";
RN Plasmid 15:93-103(1986).
RP [4]
RP REVISIONS.
RC PLASMID=pub110;
RX MEDLINE=87204663; PubMed=3033723;
RA McKenzie T., Hoshino T., Tanaka T., Sueoka N.;
RT "Correction. A revision of the nucleotide sequence and functional map
RL of pub110.";
RN Plasmid 17:83-85(1987).
RT -1- FUNCTION: THE INTERACTION OF THE RSA SITE AND THE PRE PROTEIN MAY
CC NOT ONLY SERVES A FUNCTION IN PLASMID MAINTENANCE, BUT MAY ALSO
CC CONTRIBUTE TO THE DISTRIBUTION OF SMALL ANTIBIOTIC RESISTANCE
CC PLASMIDS AMONG GRAM-POSITIVE BACTERIA.
CC -1- MISCELLANEOUS: PRE PROTEINS CONTAIN CONSERVED POSITIVELY CHARGED
CC AMINO ACIDS PROBABLY INVOLVED IN THE BINDING OF THE PRE PROTEIN TO
CC THE RSA SITE.
CC -1- SIMILARITY: TO OTHER PRE PROTEINS (FROM PLASMIDS PUB110, PMV158,
CC PE194, PT181, PTB913), IN THEIR N-TERMINAL ONLY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC
CC EMBL; AP003358; BAB56193.1; -
CC EMBL; AP003129; BAB41247.1; -
CC EMBL; M37273; AAA98212.1; -
CC EMBL; M19465; AAA88359.1; -
CC PIR; T44132; T44132.
CC InterPro: IPR001668; Mob_Pre.
CC Pfam; PF01076; Mob_Pre; 1.
CC KMW Plasmid; DNA-binding; Complete proteome.
CC FT BINDING 44 44 DNA (POTENTIAL).
CC FT BINDING 114 114 DNA (POTENTIAL).
CC SEQUENCE 420 AA; 49660 MW; 76B16BD3A2B96894 CRC64;
SQ

```

```

OK NCBI_TaxID=10401;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=Alburn 1;
RA MEDLINE=92087490; PubMed=1727613;
RN Davidson A.J.;
RT "Channel catfish virus: a new type of herpesvirus.";
RN Virology 186:9-14(1992).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC
CC EMBL; M75136; AAA88159.1; -
CC PIR; C36792; C36792.
CC KW Hypothetical protein.
SQ SEQUENCE 1179 AA; 132388 MW; B732E046AF0D8218 CRC64;

```

Query Match 56.7%; Score 38; DB 1; Length 1179;  
Best Local Similarity 45.5%; Pred. No. 1.1e+02;  
Matches 5; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

```

OY 2 ITRMDKSVNDY 12
DB 583 VTQWDATANKY 593

```

RESULT 10

```

ID PAH2_BACHD STANDARD; PRT; 332 AA.
AC O9KRT5;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE 3-oxoacyl-[acyl-carrier-protein] synthase III protein 2 (EC 2.3.1.41)
DE (Beta-oxoacyl-ACP synthase III 2) (KAS III 2).
GN PAH2 OR BH0764.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=86665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C-125 / JCM 9153;
RX MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kihara S.,
RA Horiuchi K.;
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
RT halodurans and genomic sequence comparison with Bacillus subtilis.";
RL Nucleic Acids Res. 28:4317-4331(2000).
CC -1- FUNCTION: Catalyzes the condensation reaction of fatty acid
CC synthesis by the addition to an acyl acceptor of two carbons from
CC malonyl-ACP. Catalyzes the first condensation reaction which
CC initiates fatty acid synthesis and may therefore play a role in
CC governing the total rate of fatty acid production. Possesses both
CC acetoacetyl-ACP synthase and acetyl transacylase activities. Its
CC substrate specificity determines the biosynthesis of branched-
CC chain and/or straight-chain of fatty acids (By similarity).
CC -1- CATALYTIC ACTIVITY: Acyl-[acyl-carrier protein] + malonyl-[acyl-
CC carrier protein] = 3-oxoacyl-[acyl-carrier protein] + CO(2) +
CC [acyl-carrier protein].
CC -1- PATHWAY: Fatty acid biosynthesis.
CC -1- SUBUNIT: Homodimer (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -1- DOMAIN: The last Arg residue of the ACP-binding site is essential
CC for the weak association between ACP/acp and fabH (By
CC similarity).
CC -1- SIMILARITY: Belongs to the fabH family.
CC

```

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL, AP001509; BAB04483.1; -  
 DR PIR; D83745; D83745.  
 DR HSSP; P24249; 1HNK.  
 DR HAMAP; MF\_01815; -; 1.  
 DR InterPro; IPR004655; Fadh.  
 DR TIGRFAMs; TIGR00747; Fadh; 1.  
 KM Fatty acid biosynthesis; Transferase; Acyltransferase;  
 KM Multifunctional enzyme; Complete proteome.  
 FT ACT\_SITE 115 115 BY SIMILARITY.  
 FT ACT\_SITE 252 252 BY SIMILARITY.  
 FT ACT\_SITE 282 282 BY SIMILARITY.  
 FT SITE 253 257 ACP-BINDING (BY SIMILARITY).  
 SQ SEQUENCE 332 AA; 36264 MW; DDD70C10F8B5C6D CRC64;

Query Match 55.2%; Score 37; DB 1; Length 332;  
 Best Local Similarity 63.6%; Pred. No. 44;  
 Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 LITRMDKSVND 11  
 |||:|:|:|:  
 Db 65 LIARYKSVND 75

RESULT 11  
 SYH\_MYCBE STANDARD; PRT; 414 AA.  
 AC P47281;

DT 01-FEB-1996 (Rel. 33, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Histidyl-tRNA synthetase (EC 6.1.1.21) (Histidine--tRNA ligase) (HisR).  
 DE HIS OR MG035.  
 GN Mycoplasma genitalium.  
 OS Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.  
 CC NCBI\_TaxID=2997;  
 RX NCBI\_TaxID=2997;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 33530 / G-37;  
 RX MEDLINE=96026346; PubMed=7569993;  
 RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A., Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M., Fritchman J., Weidman J.F., Small K.V., Sandusky M., Fuhmann J.L., Nguyen D.T., Uterback T.R., Saudek D.M., Phillips C.A., Merrick J.M., Tomb J.-F., Dougherty B.A., Bott K.F., Hu P.-C., Luster T.S., Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;  
 RA "The minimal gene complement of Mycoplasma genitalium.";  
 RT Science 270:397-403(1995)

CC -1 CATALYTIC ACTIVITY: ATP + L-histidine + tRNA(His) = AMP + diphosphate + L-histidyl-tRNA(His).  
 CC -1 SUBUNIT: Homodimer (By similarity).  
 CC -1 SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -1 SIMILARITY: Belongs to class-II aminoacyl-tRNA synthetase family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL, U39683; AAC71251.1; -  
 DR PIR; H64203; H64203.  
 DR HSSP; O32422; IOE0.

DR TIGR; MG035; -  
 DR HAMAP; MF\_00127; -; 1.  
 DR InterPro; IPR004516; HisS.  
 DR InterPro; IPR002314; tRNA-synt\_2b.  
 DR InterPro; IPR006195; tRNA\_ligase\_II.  
 DR Pfam; PF00587; tRNA-synt\_2b; 1.  
 DR TIGRFAMs; TIGR00442; hisS; 1.  
 DR PROSITE; PS00862; AA-tRNA\_LIGASE\_II; 1.  
 KM Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding; Complete proteome.  
 SQ SEQUENCE 414 AA; 48324 MW; 9419F5BDB12110A CRC64;

Query Match 55.2%; Score 37; DB 1; Length 414;  
 Best Local Similarity 36.4%; Pred. No. 55;  
 Matches 4; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

QY 2 ITRMDKSVNDY 12  
 ::|:|:|:|:  
 Db 173 LSKMKALKXDY 183

RESULT 12  
 ANFX\_RHOCA STANDARD; PRT; 460 AA.  
 ID ANFX\_RHOCA  
 AC 007935;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE Nitrogenase iron-iron protein beta chain (EC 1.18.6.1) (Nitrogenase component I) (Dinitrogenase 3 beta subunit).  
 GN ANFX.  
 OS Rhodospirillum rubrum.  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhodospirillales;  
 OC Rhodospirillaceae; Rhodospirillum.  
 CC NCBI\_TaxID=1061;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=B10S;  
 RX MEDLINE=93323746; PubMed=8332060;  
 RA Schmeddekopf K., Henneke S., Liese U., Kutsche M., Klipp W.;  
 RT "Characterization of ant genes specific for the alternative nitrogenase and identification of nit genes required for both nitrogenases in Rhodospirillum rubrum.";  
 RT Mol. Microbiol. 8:673-684(1993).

CC -1 FUNCTION: THE KEY ENZYMAIC REACTIONS IN NITROGEN FIXATION ARE CATALYZED BY THE NITROGENASE COMPLEX, WHICH HAS 2 COMPONENTS: THE IRON PROTEIN (COMPONENT 2) AND A COMPONENT 1 WHICH IS EITHER A MOLEBDENUM-IRON PROTEIN, A VANADIUM-IRON, OR AN IRON-IRON PROTEIN.  
 CC -1 CATALYTIC ACTIVITY: 8 reduced ferredoxin + 8 H(+) + N(2) + 16 ATP = 8 oxidized ferredoxin + 2 NH(3) + 16 ADP + 16 phosphate.  
 CC -1 COFACTOR: Iron-sulfur.  
 CC -1 SUBUNIT: HEXAMER OF TWO ALPHA, TWO BETA, AND TWO DELTA CHAINS.  
 CC -1 SIMILARITY: BELONGS TO THE NIFD/NIFE/NIFE/NIFEN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL, X70033; CAA49627.1; -  
 DR PIR; S34947; S34947.  
 DR HSSP; P11347; 1MO.  
 DR InterPro; IPR000318; Nitrogenase comp.  
 DR InterPro; IPR000510; Oxidized nitrogn.  
 DR Pfam; PF00148; oxidized nitro; 1.  
 DR PROSITE; PS00699; NITROGENASE\_1; 1.  
 DR PROSITE; PS00900; NITROGENASE\_1\_2; 1.  
 KM Oxidoreductase; Nitrogen fixation; Iron-sulfur.  
 SQ SEQUENCE 460 AA; 50702 MW; D20D361C76DAF8D4 CRC64;

Query Match 55.2%; Score 37; DB 1; Length 460;  
 Best Local Similarity 50.0%; Pred. No. 61;  
 Matches 6; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 1 LITRMDKSVNDY 12  
 :|||:|:|:  
 Db 148 MITGDKAVHDF 159

RESULT 13  
 YB89\_YEAST STANDARD; PRT; 529 AA.  
 AC P38140;  
 DT 01-OCT-1994 (Rel. 30, Created)  
 DT 01-OCT-1994 (Rel. 30, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Putative 60.3 kDa transcriptional regulatory protein in PRP5-TH12  
 DE Intergenic region.  
 DE YBR239C OR YBR1622.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OS Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 CC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 CC NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=6288c;  
 RA Ajjivovic G., Pohl F.M., Pohl T.M.;  
 RL Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.  
 CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).  
 CC -1- SIMILARITY: Contains 1 Zn(2)-Cys(6) fungal-type binuclear cluster  
 CC domain.  
 CC -1- SIMILARITY: Contains 1 PAS (PER-ARNT-SIM) dimerization domain.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC at the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; Z36108; CAAS202.1; -  
 CC PIR; S46116; S46116.  
 CC GeneMonline; 138782; -  
 CC SGP; S0000443; YBR239C.  
 CC InterPro; IPR001138; Fungi\_TN.  
 CC InterPro; IPR000014; PAS\_domain.  
 CC Pfam; PF00989; PAS; 1.  
 CC Pfam; PF00172; Zn\_c1us; 1.  
 CC SMART; SM00066; GAL4; 1.  
 CC SMART; SM00091; PAS; 1.  
 CC PROSITE; PS00463; ZN2\_Cy6\_FUNGAL\_1; 1.  
 CC PROSITE; PS50048; ZN2\_Cy6\_FUNGAL\_2; 1.  
 CC PROSITE; PS50112; PAS; 1.  
 CC Hypothetical protein; Transcription regulation; DNA-binding;  
 KW Nuclear protein; zinc; Metal-binding.  
 FT DNA BIND 408 480 PAS.  
 FT DOMAIN 408 480 PAS.  
 FT SEQUENCE 529 AA; 60310 MW; 6C99B70C8CA93D8 CRC64;

Query Match 55.2%; Score 37; DB 1; Length 529;  
 Best Local Similarity 62.5%; Pred. No. 71;  
 Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 5 WDKSVNDY 12  
 :|||:|:|:  
 Db 279 WDSINQY 286

RESULT 14  
 PEX\_HUMAN STANDARD; PRT; 749 AA.  
 ID PEX\_HUMAN  
 AC F78562; 000678; Q13646; Q93032; Q99827;

DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Phosphite regulating neutral endopeptidase (EC 3.4.24.-)  
 DE (Metalloendopeptidase homolog PEX) (X-linked hypophosphatemia protein)  
 DE (HYP) (Vitamin D-resistant hypophosphatemic rickets protein).  
 GN PHEX OR PEX.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 CC NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND VARIANTS HYP ARG-85; LEU-534; ARG-579 AND  
 RP PRO-651.  
 RX MEDLINE=97343325; PubMed=9199930;  
 RA Francis F., Strom T.M., Hennig S., Boeddrich A., Lorenz B.,  
 RA Brandau O., Mohnik K.L., Cagnoli M., Steffens C., Klages S.,  
 RA Borzym K., Pohl T., Oudet C.L., Rowe P.S.N., Reinhardt R.,  
 RA Meitinger T., Leirach H.;  
 RT "Genomic organization of the human PEX gene mutated in X-linked  
 RT dominant hypophosphatemic rickets.";  
 RT Genome Res. 7:573-585 (1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97232252; PubMed=9077527;  
 RA Beck L., Soumounou Y., Martel J., Krishnamurthy G., Gauthier C.,  
 RA Goodyer C.G., Tenenhouse H.S.;  
 RT "PeX/PEX tissue distribution and evidence for a deletion in the 3'  
 RT region of the Pex gene in X-linked hypophosphatemic mice.";  
 RL J. Clin. Invest. 99:1200-1209 (1997).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Bone;  
 RX MEDLINE=97343443; PubMed=9199999;  
 RA Guo R., Quarles L.D.;  
 RT "Cloning and sequencing of human PEX from a bone cDNA library:  
 RT evidence for its developmental stage-specific regulation in  
 RT osteoblasts.";  
 RL J. Bone Miner. Res. 12:1009-1017 (1997).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA Lipman M.L., Panda D., Henderson J.E., Shen Y., Goltzman D.,  
 RA Karaplis A.C.;  
 RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97224400; PubMed=9070861;  
 RA Grieff M., Mumm S., Maeltz P., Mazzarella R., Whyte M.P.,  
 RA Trakker R.V., Schlessinger D.;  
 RT "Expression and cloning of the human X-linked hypophosphatemia gene  
 RT cDNA.";  
 RL Biochem. Biophys. Res. Commun. 231:635-639 (1997).  
 RN [6]  
 RP SEQUENCE OF 4-641 FROM N.A.  
 RX MEDLINE=96044647; PubMed=7550339;  
 RA Francis F., Hennig S., Korn B., Reinhardt R., de Jong P., Poustka A.,  
 RA Leirach H., Rowe P.S.N., Goulding J.N., Summerfield I., Mounford R.,  
 RA Read A.P., Popowska E., Pronicka E., Davies K.E., Orlordan J.B.H.,  
 RA Econs M.J., Nesbitt T., Dreznar M.K., Oudet C.L., Pannetier S.,  
 RA Hanauer A., Strom T.M., Meindl A., Lorenz B., Cagnoli M.,  
 RA Mohnik K.L., Murken J., Meitinger T.;  
 RT "A gene (PEX) with homologies to endopeptidases is mutated in  
 RT patients with X-linked hypophosphatemic rickets. The HYP  
 RT Consortium.";  
 RL Nat. Genet. 11:130-136 (1995).  
 RN [7]  
 RP SEQUENCE OF 1-116 FROM N.A.  
 RP Waterston R.;  
 RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.  
 RN [8]  
 RP VARIANTS HYP TYR-85; CYS-166; SER-252; ILE-253 AND VAL-579.  
 RX MEDLINE=97260404; PubMed=9106524;  
 RA Holm I.A., Huang X., Kunkel L.M.;

RT "Mutational analysis of the PEX gene in patients with X-linked  
RT hypophosphatemic rickets.";  
RL Am. J. Hum. Genet. 60:790-797(1997).  
RN [9]  
RP VARIANTS HYP SER-77; PRO-138; LEU-534 AND ARG-579.  
RX MEDLINE=97252387; PubMed=9097956;  
RA Rowe P.S.N., Oudet C.L., Francis F., Sindling C., Pannetier S.,  
RA Econs M.J., Strom T.M., Melinger T., Garabedian M., David A.,  
RA Macher M.-A., Questiaux E., Popowska E., Pronicka E., Read A.P.,  
RA Mokrycki A., Glorieux F.H., Dreiner M.K., Hanauer A., Lebrach H.,  
RA Goulding U.N., O'Riordan J.L.H.;  
RT "Distribution of mutations in the PEX gene in families with X-linked  
RT hypophosphatemic rickets (HYP)."  
RL Hum. Mol. Genet. 6:539-549(1997).  
RN [10]  
RP VARIANT HYP PRO-555.  
RX MEDLINE=94439582; PubMed=9768646;  
RA Econs M.J., Friedman N.B., Rowe P.S.N., Speer M.C., Francis F.,  
RA Strom T.M., Oudet C.L., Smith J.A., Ninomiya U.T., Lee B.E.,  
RA Bergen H.;  
RT "A PHEX gene mutation is responsible for adult-onset vitamin  
RT D-resistant hypophosphatemic osteomalacia: evidence that the disorder  
RT is not a distinct entity from X-linked hypophosphatemic rickets.";  
RL J. Clin. Endocrinol. Metab. 83:3459-3462(1998).  
RN [11]  
RP VARIANTS HYP PHE-317; LEU-534; ARG-579; ARG-621; ASN-680 DEL; THR-720;  
RP TYR-731 AND ARG-749.  
RX MEDLINE=98439610; PubMed=9768674;  
RA Dixon P.H., Christie P.T., Wooding C., Trump D., Grief M., Holm I.A.,  
RA Gerner U.M., Schmidtke J., Shah B., Shaw N., Smith C., Tau C.,  
RA Schlessinger D., Whyte M.P., Thakker R.V.;  
RT "Mutational analysis of PHEX gene in X-linked hypophosphatemia.";  
RL J. Clin. Endocrinol. Metab. 83:3615-3623(1998).  
RN [12]  
RP VARIANTS HYP SER-80; PHE-142; GLY-237; CYS-530; ASP-573; SER-733 AND  
RP TRP-746.  
RX MEDLINE=98368844; PubMed=10439971.  
RA Filisetti D., Ostermann G., von Bredow M., Strom T.M., Filler G.,  
RA Ehrlich J., Pannetier S., Gerner U.M., Rowe P.S.N., Francis F.,  
RA Juulien A., Hanauer A., Econs M.J., Oudet C.L.;  
RT "Non-random distribution of mutations in the PHEX gene, and  
RT under-detected missense mutations at non-conserved residues.";  
RL Eur. J. Hum. Genet. 7:615-619(1999).  
RN [13]  
RP VARIANTS HYP PHE-85; PRO-141; VAL-341 DEL; PRO-567; LYS-680 AND  
RP TYR-693.  
RX MEDLINE=20202840; PubMed=10737991;  
RA Tyymismaa H., Kaitila I., Naentoe-Salonen K., Ala-Houhaja M.,  
RA Alitalo T.;  
RT "Identification of fifteen novel PHEX gene mutations in Finnish  
RT patients with hypophosphatemic rickets.";  
RL Hum. Mutat. 15:383-384(2000).  
RN [14]  
RP VARIANTS HYP ARG-160 AND ASN-444 INS.  
RX MEDLINE=20461419; PubMed=11004247;  
RA Sato K., Tajima T., Nakae J., Adachi M., Asakura Y., Tachibana K.,  
RA Suwa S., Katsumata N., Tanaka T., Hayashi Y., Abe S., Morishita M.,  
RA Okuhara K., Shinohara N., Fujieda K.;  
RT "Three novel PHEX gene mutations in Japanese patients with X-linked  
RT hypophosphatemic rickets.";  
RL Pediatr. Res. 48:536-540(2000).  
RN [15]  
CC -1- FUNCTION: Probably involved in bone and dentin mineralization  
CC and renal phosphate reabsorption.  
CC -1- COFACTOR: Binds 1 zinc ion per subunit (By similarity).  
CC -1- SUBCELLULAR LOCATION: Type II membrane protein (Potential).  
CC -1- TISSUE SPECIFICITY: Lymphocyte and fetal brain; not in adult  
CC brain, placenta, skeletal muscle and pancreas; not in adult and  
CC fetal heart, lung, liver and kidney.  
CC -1- DISEASE: Defects in PHEX are a cause of X-linked hypophosphatemic  
CC rickets (HYP) [MIM:307800]. HYP is an X-linked dominant disorder  
CC characterized by impaired phosphate uptake in the kidney, which is  
CC likely to be caused by abnormal regulation of sodium phosphate  
CC cotransport in the proximal tubules. Clinical manifestations

CC include skeletal deformities, growth failure, craniosynostosis,  
CC paravertebral calcifications, pseudofractures in lower  
CC extremities, and muscular hypotonia with onset in early childhood.  
CC X-linked hypophosphatemic rickets is the most common form of  
CC hypophosphatemia with an incidence of 1 in 20000.  
CC -1- SIMILARITY: Belongs to peptidase family M13.  
CC -1- DATABASE: NAME=PEXdb; WWW=<http://data.mch.mcgill.ca/phexdb/>.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; Y10196; CAA71258.1; -  
DR EMBL; U75645; AAB47749.1; -  
DR EMBL; U82970; AAC24487.1; -  
DR EMBL; U87284; AAB47562.1; -  
DR EMBL; Y08111; CAA69326.1; -  
DR EMBL; Y08113; CAA69326.1; JOINED.  
DR EMBL; Y08113; CAA69326.1; JOINED.  
DR EMBL; Y08114; CAA69326.1; JOINED.  
DR EMBL; Y08115; CAA69326.1; JOINED.  
DR EMBL; Y08116; CAA69326.1; JOINED.  
DR EMBL; Y08117; CAA69326.1; JOINED.  
DR EMBL; Y08118; CAA69326.1; JOINED.  
DR EMBL; Y08119; CAA69326.1; JOINED.  
DR EMBL; Y08120; CAA69326.1; JOINED.  
DR EMBL; Y08121; CAA69326.1; JOINED.  
DR EMBL; Y08122; CAA69326.1; JOINED.  
DR EMBL; Y08123; CAA69326.1; JOINED.  
DR EMBL; Y08124; CAA69326.1; JOINED.  
DR EMBL; Y08125; CAA69326.1; JOINED.  
DR EMBL; Y08126; CAA69326.1; JOINED.  
DR EMBL; Y08127; CAA69326.1; JOINED.  
DR EMBL; Y08128; CAA69326.1; JOINED.  
DR EMBL; Y08129; CAA69326.1; JOINED.  
DR EMBL; Y08130; CAA69326.1; JOINED.  
DR EMBL; Y08131; CAA69326.1; JOINED.  
DR EMBL; Y08132; CAA69326.1; JOINED.  
DR EMBL; Y08133; CAA69326.1; JOINED.  
DR EMBL; Y08134; AAB42219.1; -  
DR EMBL; Y08135; AAB42219.1; JOINED.  
DR EMBL; Y08136; AAB42219.1; JOINED.  
DR EMBL; Y08137; AAB42219.1; JOINED.  
DR EMBL; Y08138; AAB42219.1; JOINED.  
DR EMBL; Y08139; AAB42219.1; JOINED.  
DR EMBL; Y08140; AAB42219.1; JOINED.  
DR EMBL; Y08141; AAB42219.1; JOINED.  
DR EMBL; Y08142; AAB42219.1; JOINED.  
DR EMBL; Y08143; AAB42219.1; JOINED.  
DR EMBL; Y08144; AAB42219.1; JOINED.  
DR EMBL; Y08145; AAB42219.1; JOINED.  
DR EMBL; Y08146; AAB42219.1; JOINED.  
DR EMBL; Y08147; AAB42219.1; JOINED.  
DR EMBL; Y08148; AAB42219.1; JOINED.  
DR EMBL; Y08149; AAB42219.1; JOINED.  
DR EMBL; Y08150; AAB42219.1; JOINED.  
DR EMBL; Y08151; AAB42219.1; JOINED.  
DR EMBL; Y08152; AAB42219.1; JOINED.  
DR EMBL; Y08153; AAB42219.1; JOINED.  
DR EMBL; Y08154; AAB42219.1; JOINED.  
DR EMBL; Y08155; AAB42219.1; JOINED.  
DR EMBL; Y08156; AAB42219.1; JOINED.  
DR EMBL; Y08157; AAB42219.1; JOINED.  
DR EMBL; Y08158; AAB42219.1; JOINED.  
DR EMBL; Y08159; AAB42219.1; JOINED.  
DR EMBL; Y08160; AAB42219.1; JOINED.  
DR EMBL; Y08161; AAB42219.1; JOINED.  
DR EMBL; Y08162; AAB42219.1; JOINED.  
DR EMBL; Y08163; AAB42219.1; JOINED.  
DR EMBL; Y08164; AAB42219.1; JOINED.  
DR EMBL; Y08165; AAB42219.1; JOINED.  
DR EMBL; Y08166; AAB42219.1; JOINED.  
DR EMBL; Y08167; AAB42219.1; JOINED.  
DR EMBL; Y08168; AAB42219.1; JOINED.  
DR EMBL; Y08169; AAB42219.1; JOINED.  
DR EMBL; Y08170; AAB42219.1; JOINED.  
DR EMBL; Y08171; AAB42219.1; JOINED.  
DR EMBL; Y08172; AAB42219.1; JOINED.  
DR EMBL; Y08173; AAB42219.1; JOINED.  
DR EMBL; Y08174; AAB42219.1; JOINED.  
DR EMBL; Y08175; AAB42219.1; JOINED.  
DR EMBL; Y08176; AAB42219.1; JOINED.  
DR EMBL; Y08177; AAB42219.1; JOINED.  
DR EMBL; Y08178; AAB42219.1; JOINED.  
DR EMBL; Y08179; AAB42219.1; JOINED.  
DR EMBL; Y08180; AAB42219.1; JOINED.  
DR EMBL; Y08181; AAB42219.1; JOINED.  
DR EMBL; Y08182; AAB42219.1; JOINED.  
DR EMBL; Y08183; AAB42219.1; JOINED.  
DR EMBL; Y08184; AAB42219.1; JOINED.  
DR EMBL; Y08185; AAB42219.1; JOINED.  
DR EMBL; Y08186; AAB42219.1; JOINED.  
DR EMBL; Y08187; AAB42219.1; JOINED.  
DR EMBL; Y08188; AAB42219.1; JOINED.  
DR EMBL; Y08189; AAB42219.1; JOINED.  
DR EMBL; Y08190; AAB42219.1; JOINED.  
DR EMBL; Y08191; AAB42219.1; JOINED.  
DR EMBL; Y08192; AAB42219.1; JOINED.  
DR EMBL; Y08193; AAB42219.1; JOINED.  
DR EMBL; Y08194; AAB42219.1; JOINED.  
DR EMBL; Y08195; AAB42219.1; JOINED.  
DR EMBL; Y08196; AAB42219.1; JOINED.  
DR EMBL; Y08197; AAB42219.1; JOINED.  
DR EMBL; Y08198; AAB42219.1; JOINED.  
DR EMBL; Y08199; AAB42219.1; JOINED.  
DR EMBL; Y08200; AAB42219.1; JOINED.  
DR EMBL; Y08201; AAB42219.1; JOINED.  
DR EMBL; Y08202; AAB42219.1; JOINED.  
DR EMBL; Y08203; AAB42219.1; JOINED.  
DR EMBL; Y08204; AAB42219.1; JOINED.  
DR EMBL; Y08205; AAB42219.1; JOINED.  
DR EMBL; Y08206; AAB42219.1; JOINED.  
DR EMBL; Y08207; AAB42219.1; JOINED.  
DR EMBL; Y08208; AAB42219.1; JOINED.  
DR EMBL; Y08209; AAB42219.1; JOINED.  
DR EMBL; Y08210; AAB42219.1; JOINED.  
DR EMBL; Y08211; AAB42219.1; JOINED.  
DR EMBL; Y08212; AAB42219.1; JOINED.  
DR EMBL; Y08213; AAB42219.1; JOINED.  
DR EMBL; Y08214; AAB42219.1; JOINED.  
DR EMBL; Y08215; AAB42219.1; JOINED.  
DR EMBL; Y08216; AAB42219.1; JOINED.  
DR EMBL; Y08217; AAB42219.1; JOINED.  
DR EMBL; Y08218; AAB42219.1; JOINED.  
DR EMBL; Y08219; AAB42219.1; JOINED.  
DR EMBL; Y08220; AAB42219.1; JOINED.  
DR EMBL; Y08221; AAB42219.1; JOINED.  
DR EMBL; Y08222; AAB42219.1; JOINED.  
DR EMBL; Y08223; AAB42219.1; JOINED.  
DR EMBL; Y08224; AAB42219.1; JOINED.  
DR EMBL; Y08225; AAB42219.1; JOINED.  
DR EMBL; Y08226; AAB42219.1; JOINED.  
DR EMBL; Y08227; AAB42219.1; JOINED.  
DR EMBL; Y08228; AAB42219.1; JOINED.  
DR EMBL; Y08229; AAB42219.1; JOINED.  
DR EMBL; Y08230; AAB42219.1; JOINED.  
DR EMBL; Y08231; AAB42219.1; JOINED.  
DR EMBL; Y08232; AAB42219.1; JOINED.  
DR EMBL; Y08233; AAB42219.1; JOINED.  
DR EMBL; Y08234; AAB42219.1; JOINED.  
DR EMBL; Y08235; AAB42219.1; JOINED.  
DR EMBL; Y08236; AAB42219.1; JOINED.  
DR EMBL; Y08237; AAB42219.1; JOINED.  
DR EMBL; Y08238; AAB42219.1; JOINED.  
DR EMBL; Y08239; AAB42219.1; JOINED.  
DR EMBL; Y08240; AAB42219.1; JOINED.  
DR EMBL; Y08241; AAB42219.1; JOINED.  
DR EMBL; Y08242; AAB42219.1; JOINED.  
DR EMBL; Y08243; AAB42219.1; JOINED.  
DR EMBL; Y08244; AAB42219.1; JOINED.  
DR EMBL; Y08245; AAB42219.1; JOINED.  
DR EMBL; Y08246; AAB42219.1; JOINED.  
DR EMBL; Y08247; AAB42219.1; JOINED.  
DR EMBL; Y08248; AAB42219.1; JOINED.  
DR EMBL; Y08249; AAB42219.1; JOINED.  
DR EMBL; Y08250; AAB42219.1; JOINED.  
DR EMBL; Y08251; AAB42219.1; JOINED.  
DR EMBL; Y08252; AAB42219.1; JOINED.  
DR EMBL; Y08253; AAB42219.1; JOINED.  
DR EMBL; Y08254; AAB42219.1; JOINED.  
DR EMBL; Y08255; AAB42219.1; JOINED.  
DR EMBL; Y08256; AAB42219.1; JOINED.  
DR EMBL; Y08257; AAB42219.1; JOINED.  
DR EMBL; Y08258; AAB42219.1; JOINED.  
DR EMBL; Y08259; AAB42219.1; JOINED.  
DR EMBL; Y08260; AAB42219.1; JOINED.  
DR EMBL; Y08261; AAB42219.1; JOINED.  
DR EMBL; Y08262; AAB42219.1; JOINED.  
DR EMBL; Y08263; AAB42219.1; JOINED.  
DR EMBL; Y08264; AAB42219.1; JOINED.  
DR EMBL; Y08265; AAB42219.1; JOINED.  
DR EMBL; Y08266; AAB42219.1; JOINED.  
DR EMBL; Y08267; AAB42219.1; JOINED.  
DR EMBL; Y08268; AAB42219.1; JOINED.  
DR EMBL; Y08269; AAB42219.1; JOINED.  
DR EMBL; Y08270; AAB42219.1; JOINED.  
DR EMBL; Y08271; AAB42219.1; JOINED.  
DR EMBL; Y08272; AAB42219.1; JOINED.  
DR EMBL; Y08273; AAB42219.1; JOINED.  
DR EMBL; Y08274; AAB42219.1; JOINED.  
DR EMBL; Y08275; AAB42219.1; JOINED.  
DR EMBL; Y08276; AAB42219.1; JOINED.  
DR EMBL; Y08277; AAB42219.1; JOINED.  
DR EMBL; Y08278; AAB42219.1; JOINED.  
DR EMBL; Y08279; AAB42219.1; JOINED.  
DR EMBL; Y08280; AAB42219.1; JOINED.  
DR EMBL; Y08281; AAB42219.1; JOINED.  
DR EMBL; Y08282; AAB42219.1; JOINED.  
DR EMBL; Y08283; AAB42219.1; JOINED.  
DR EMBL; Y08284; AAB42219.1; JOINED.  
DR EMBL; Y08285; AAB42219.1; JOINED.  
DR EMBL; Y08286; AAB42219.1; JOINED.  
DR EMBL; Y08287; AAB42219.1; JOINED.  
DR EMBL; Y08288; AAB42219.1; JOINED.  
DR EMBL; Y08289; AAB42219.1; JOINED.  
DR EMBL; Y08290; AAB42219.1; JOINED.  
DR EMBL; Y08291; AAB42219.1; JOINED.  
DR EMBL; Y08292; AAB42219.1; JOINED.  
DR EMBL; Y08293; AAB42219.1; JOINED.  
DR EMBL; Y08294; AAB42219.1; JOINED.  
DR EMBL; Y08295; AAB42219.1; JOINED.  
DR EMBL; Y08296; AAB42219.1; JOINED.  
DR EMBL; Y08297; AAB42219.1; JOINED.  
DR EMBL; Y08298; AAB42219.1; JOINED.  
DR EMBL; Y08299; AAB42219.1; JOINED.  
DR EMBL; Y08300; AAB42219.1; JOINED.  
DR EMBL; Y08301; AAB42219.1; JOINED.  
DR EMBL; Y08302; AAB42219.1; JOINED.  
DR EMBL; Y08303; AAB42219.1; JOINED.  
DR EMBL; Y08304; AAB42219.1; JOINED.  
DR EMBL; Y08305; AAB42219.1; JOINED.  
DR EMBL; Y08306; AAB42219.1; JOINED.  
DR EMBL; Y08307; AAB42219.1; JOINED.  
DR EMBL; Y08308; AAB42219.1; JOINED.  
DR EMBL; Y08309; AAB42219.1; JOINED.  
DR EMBL; Y08310; AAB42219.1; JOINED.  
DR EMBL; Y08311; AAB42219.1; JOINED.  
DR EMBL; Y08312; AAB42219.1; JOINED.  
DR EMBL; Y08313; AAB42219.1; JOINED.  
DR EMBL; Y08314; AAB42219.1; JOINED.  
DR EMBL; Y08315; AAB42219.1; JOINED.  
DR EMBL; Y08316; AAB42219.1; JOINED.  
DR EMBL; Y08317; AAB42219.1; JOINED.  
DR EMBL; Y08318; AAB42219.1; JOINED.  
DR EMBL; Y08319; AAB42219.1; JOINED.  
DR EMBL; Y08320; AAB42219.1; JOINED.  
DR EMBL; Y08321; AAB42219.1; JOINED.  
DR EMBL; Y08322; AAB42219.1; JOINED.  
DR EMBL; Y08323; AAB42219.1; JOINED.  
DR EMBL; Y08324; AAB42219.1; JOINED.  
DR EMBL; Y08325; AAB42219.1; JOINED.  
DR EMBL; Y08326; AAB42219.1; JOINED.  
DR EMBL; Y08327; AAB42219.1; JOINED.  
DR EMBL; Y08328; AAB42219.1; JOINED.  
DR EMBL; Y08329; AAB42219.1; JOINED.  
DR EMBL; Y08330; AAB42219.1; JOINED.  
DR EMBL; Y08331; AAB42219.1; JOINED.  
DR EMBL; Y08332; AAB42219.1; JOINED.  
DR EMBL; Y08333; AAB42219.1; JOINED.  
DR EMBL; Y08334; AAB42219.1; JOINED.  
DR EMBL; Y08335; AAB42219.1; JOINED.  
DR EMBL; Y08336; AAB42219.1; JOINED.  
DR EMBL; Y08337; AAB42219.1; JOINED.  
DR EMBL; Y08338; AAB42219.1; JOINED.  
DR EMBL; Y08339; AAB42219.1; JOINED.  
DR EMBL; Y08340; AAB42219.1; JOINED.  
DR EMBL; Y08341; AAB42219.1; JOINED.  
DR EMBL; Y08342; AAB42219.1; JOINED.  
DR EMBL; Y08343; AAB42219.1; JOINED.  
DR EMBL; Y08344; AAB42219.1; JOINED.  
DR EMBL; Y08345; AAB42219.1; JOINED.  
DR EMBL; Y08346; AAB42219.1; JOINED.  
DR EMBL; Y08347; AAB42219.1; JOINED.  
DR EMBL; Y08348; AAB42219.1; JOINED.  
DR EMBL; Y08349; AAB42219.1; JOINED.  
DR EMBL; Y08350; AAB42219.1; JOINED.  
DR EMBL; Y08351; AAB42219.1; JOINED.  
DR EMBL; Y08352; AAB42219.1; JOINED.  
DR EMBL; Y08353; AAB42219.1; JOINED.  
DR EMBL; Y08354; AAB42219.1; JOINED.  
DR EMBL; Y08355; AAB42219.1; JOINED.  
DR EMBL; Y08356; AAB42219.1; JOINED.  
DR EMBL; Y08357; AAB42219.1; JOINED.  
DR EMBL; Y08358; AAB42219.1; JOINED.  
DR EMBL; Y08359; AAB42219.1; JOINED.  
DR EMBL; Y08360; AAB42219.1; JOINED.  
DR EMBL; Y08361; AAB42219.1; JOINED.  
DR EMBL; Y08362; AAB42219.1; JOINED.  
DR EMBL; Y08363; AAB42219.1; JOINED.  
DR EMBL; Y08364; AAB42219.1; JOINED.  
DR EMBL; Y08365; AAB42219.1; JOINED.  
DR EMBL; Y08366; AAB42219.1; JOINED.  
DR EMBL; Y08367; AAB42219.1; JOINED.  
DR EMBL; Y08368; AAB42219.1; JOINED.  
DR EMBL; Y08369; AAB42219.1; JOINED.  
DR EMBL; Y08370; AAB42219.1; JOINED.  
DR EMBL; Y08371; AAB42219.1; JOINED.  
DR EMBL; Y08372; AAB42219.1; JOINED.  
DR EMBL; Y08373; AAB42219.1; JOINED.  
DR EMBL; Y08374; AAB42219.1; JOINED.  
DR EMBL; Y08375; AAB42219.1; JOINED.  
DR EMBL; Y08376; AAB42219.1; JOINED.  
DR EMBL; Y08377; AAB42219.1; JOINED.  
DR EMBL; Y08378; AAB42219.1; JOINED.  
DR EMBL; Y08379; AAB42219.1; JOINED.  
DR EMBL; Y08380; AAB42219.1; JOINED.  
DR EMBL; Y08381; AAB42219.1; JOINED.  
DR EMBL; Y08382; AAB42219.1; JOINED.  
DR EMBL; Y08383; AAB42219.1; JOINED.  
DR EMBL; Y08384; AAB42219.1; JOINED.  
DR EMBL; Y08385; AAB42219.1; JOINED.  
DR EMBL; Y08386; AAB42219.1; JOINED.  
DR EMBL; Y08387; AAB42219.1; JOINED.  
DR EMBL; Y08388; AAB42219.1; JOINED.  
DR EMBL; Y08389; AAB42219.1; JOINED.  
DR EMBL; Y08390; AAB42219.1; JOINED.  
DR EMBL; Y08391; AAB42219.1; JOINED.  
DR EMBL; Y08392; AAB42219.1; JOINED.  
DR EMBL; Y08393; AAB42219.1; JOINED.  
DR EMBL; Y08394; AAB42219.1; JOINED.  
DR EMBL; Y08395; AAB42219.1; JOINED.  
DR EMBL; Y08396; AAB42219.1; JOINED.  
DR EMBL; Y08397; AAB42219.1; JOINED.  
DR EMBL; Y08398; AAB42219.1; JOINED.  
DR EMBL; Y08399; AAB42219.1; JOINED.  
DR EMBL; Y08400; AAB42219.1; JOINED.  
DR EMBL; Y08401; AAB42219.1; JOINED.  
DR EMBL; Y08402; AAB42219.1; JOINED.  
DR EMBL; Y08403; AAB42219.1; JOINED.  
DR EMBL; Y08404; AAB42219.1; JOINED.  
DR EMBL; Y08405; AAB42219.1; JOINED.  
DR EMBL; Y08406; AAB42219.1; JOINED.  
DR EMBL; Y08407; AAB42219.1; JOINED.  
DR EMBL; Y08408; AAB42219.1; JOINED.  
DR EMBL; Y08409; AAB42219.1; JOINED.  
DR EMBL; Y08410; AAB42219.1; JOINED.  
DR EMBL; Y08411; AAB42219.1; JOINED.  
DR EMBL; Y08412; AAB42219.1; JOINED.  
DR EMBL; Y08413; AAB42219.1; JOINED.  
DR EMBL; Y08414; AAB42219.1; JOINED.  
DR EMBL; Y08415; AAB42219.1; JOINED.  
DR EMBL; Y08416; AAB42219.1; JOINED.  
DR EMBL; Y08417; AAB42219.1; JOINED.  
DR EMBL; Y08418; AAB42219.1; JOINED.  
DR EMBL; Y08419; AAB42219.1; JOINED.  
DR EMBL; Y08420; AAB42219.1; JOINED.  
DR EMBL; Y08421; AAB42219.1; JOINED.  
DR EMBL; Y08422; AAB42219.1; JOINED.  
DR EMBL; Y08423; AAB42219.1; JOINED.  
DR EMBL; Y08424; AAB42219.1; JOINED.  
DR EMBL; Y08425; AAB42219.1; JOINED.  
DR EMBL; Y08426; AAB42219.1; JOINED.  
DR EMBL; Y08427; AAB42219.1; JOINED.  
DR EMBL; Y08428; AAB42219.1; JOINED.  
DR EMBL; Y08429; AAB42219.1; JOINED.  
DR EMBL; Y08430; AAB42219.1; JOINED.  
DR EMBL; Y08431; AAB42219.1; JOINED.  
DR EMBL; Y08432; AAB42219.1; JOINED.  
DR EMBL; Y08433; AAB42219.1; JOINED.  
DR EMBL; Y08434; AAB42219.1; JOINED.  
DR EMBL; Y08435; AAB42219.1; JOINED.  
DR EMBL; Y08436; AAB42219.1; JOINED.  
DR EMBL; Y08437; AAB42219.1; JOINED.  
DR EMBL; Y08438; AAB42219.1; JOINED.  
DR EMBL; Y08439; AAB42219.1; JOINED.  
DR EMBL; Y08440; AAB42219.1; JOINED.  
DR EMBL; Y08441; AAB42219.1; JOINED.  
DR EMBL; Y08442; AAB42219.1; JOINED.  
DR EMBL; Y08443; AAB42219.1; JOINED.  
DR EMBL; Y08444; AAB42219.1; JOINED.  
DR EMBL; Y08445; AAB42219.1; JOINED.  
DR EMBL; Y08446; AAB42219.1; JOINED.  
DR EMBL; Y08447; AAB42219.1; JOINED.  
DR EMBL; Y08448; AAB42219.1; JOINED.  
DR EMBL; Y08449; AAB42219.1; JOINED.  
DR EMBL; Y08450; AAB42219.1; JOINED.  
DR EMBL; Y08451; AAB42219.1; JOINED.  
DR EMBL; Y08452; AAB42219.1; JOINED.  
DR EMBL; Y08453; AAB42219.1; JOINED.  
DR EMBL; Y08454; AAB42219.1; JOINED.  
DR EMBL; Y08455; AAB42219.1; JOINED.  
DR EMBL; Y08456; AAB42219.1; JOINED.  
DR EMBL; Y08457; AAB42219.1; JOINED.  
DR EMBL; Y08458; AAB42219.1; JOINED.  
DR EMBL; Y08459; AAB42219.1; JOINED.  
DR EMBL; Y08460; AAB42219.1; JOINED.  
DR EMBL; Y08461; AAB42219.1; JOINED.  
DR EMBL; Y08462; AAB42219.1; JOINED.  
DR EMBL; Y08463; AAB42219.1; JOINED.  
DR EMBL; Y08464; AAB42219.1; JOINED.  
DR EMBL; Y08465; AAB42219.1; JOINED.  
DR EMBL; Y08466; AAB42219.1; JOINED.  
DR EMBL; Y08467; AAB42219.1; JOINED.  
DR EMBL; Y08468; AAB42219.1; JOINED.  
DR EMBL; Y08469; AAB42219.1; JOINED.  
DR EMBL; Y08470; AAB42219.1; JOINED.  
DR EMBL; Y08471; AAB42219.1; JOINED.  
DR EMBL; Y08472; AAB42219.1; JOINED.  
DR EMBL; Y08473; AAB42219.1; JOINED.  
DR EMBL; Y08474; AAB42219.1; JOINED.  
DR EMBL; Y08475; AAB42219.1; JOINED.  
DR EMBL; Y08476; AAB42219.1; JOINED.  
DR EMBL; Y08477; AAB42219.1; JOINED.  
DR EMBL; Y08478; AAB42219.1; JOINED.  
DR EMBL; Y08479; AAB42219.1; JOINED.  
DR EMBL; Y08480; AAB42219.1; JOINED.  
DR EMBL; Y08481; AAB42219.1; JOINED.  
DR EMBL; Y08482; AAB42219.1; JOINED.  
DR EMBL; Y08483; AAB42219.1; JOINED.  
DR EMBL; Y08484; AAB42219.1; JOINED.  
DR EMBL; Y08485; AAB42219.1; JOINED.  
DR EMBL; Y08486; AAB42219.1; JOINED.  
DR EMBL; Y08487; AAB42219.1; JOINED.  
DR EMBL; Y08488; AAB42219.1; JOINED.  
DR EMBL; Y0848



This Page Blank (uspto)



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2004, 07:34:11 ; Search time 6.35185 Seconds  
(without alignments)  
596.081 Million cell updates/sec

Title: US-09-805-290A-13

Perfect score: 67

Sequence: 1 LITRMDKSVNDY 12

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 31518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

SPRMBL 25:\*

1: sp\_archaea:\*

2: sp\_bacteria:\*

3: sp\_fungi:\*

4: sp\_human:\*

5: sp\_invertebrate:\*

6: sp\_mammal:\*

7: sp\_mhc:\*

8: sp\_organelle:\*

9: sp\_phage:\*

10: sp\_plant:\*

11: sp\_protist:\*

12: sp\_virus:\*

13: sp\_vertebrate:\*

14: sp\_unclassified:\*

15: sp\_virus:\*

16: sp\_bacteriophage:\*

17: sp\_archaeal:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	47	70.1	272 5 Q9TZ13	Q9TZ13 caenorhabditis
2	47	70.1	722 2 Q9FA45	Q9FA45 streptococcus
3	42	62.7	858 15 Q9YK4	Q9YK4 human immun
4	41	61.2	77 9 Q857E5	Q857E5 mycobacterium
5	41	61.2	99 16 Q82K72	Q82K72 streptococcus
6	41	61.2	479 16 Q9K693	Q9K693 bacillus th
7	41	61.2	562 10 Q9ZRW0	Q9ZRW0 arabidopsis
8	41	61.2	865 15 Q9CQW8	Q9CQW8 human immun
9	40	59.7	76 17 Q8T1W0	Q8T1W0 methanosa
10	40	59.7	117 15 Q9YR83	Q9YR83 human immun
11	40	59.7	117 15 Q9YR82	Q9YR82 human immun
12	40	59.7	118 15 Q9ESR4	Q9ESR4 human immun
13	40	59.7	118 15 Q9ESQ8	Q9ESQ8 human immun
14	40	59.7	122 15 Q9ODQ4	Q9ODQ4 human immun
15	40	59.7	122 15 Q9ODY1	Q9ODY1 human immun
16	40	59.7	122 15 Q91LL3	Q91LL3 human immun

17	40	59.7	122 15 Q9ODN6	Q9ODN6 human immun
18	40	59.7	122 15 Q9ODP4	Q9ODP4 human immun
19	40	59.7	122 15 Q9ODK3	Q9ODK3 human immun
20	40	59.7	122 15 Q9ODW2	Q9ODW2 human immun
21	40	59.7	122 15 Q9ODP1	Q9ODP1 human immun
22	40	59.7	122 15 Q9ODK7	Q9ODK7 human immun
23	40	59.7	122 15 Q9ODJ2	Q9ODJ2 human immun
24	40	59.7	122 15 Q9ODG6	Q9ODG6 human immun
25	40	59.7	122 15 Q9ODQ3	Q9ODQ3 human immun
26	40	59.7	122 15 Q9ODR4	Q9ODR4 human immun
27	40	59.7	122 15 Q9ODR9	Q9ODR9 human immun
28	40	59.7	122 15 Q9ODR5	Q9ODR5 human immun
29	40	59.7	122 15 Q9ODQ17	Q9ODQ17 human immun
30	40	59.7	122 15 Q9ODR96	Q9ODR96 human immun
31	40	59.7	122 15 Q9ODR45	Q9ODR45 human immun
32	40	59.7	122 15 Q9ODQ07	Q9ODQ07 human immun
33	40	59.7	122 15 Q9ODQ48	Q9ODQ48 human immun
34	40	59.7	122 15 Q9ODJ48	Q9ODJ48 human immun
35	40	59.7	122 15 Q9ODC64	Q9ODC64 arabidopsis
36	40	59.7	122 15 Q9ODFX1	Q9ODFX1 glycolysis
37	40	59.7	122 15 Q9ODT4	Q9ODT4 amastax max
38	40	59.7	122 15 Q9ODT01	Q9ODT01 arabidopsis
39	40	59.7	122 8 Q9SAR2	Q9SAR2 galactose
40	40	59.7	122 16 Q9DIT06	Q9DIT06 nitrosomonas
41	40	59.7	122 15 Q9S512	Q9S512 human immun
42	40	59.7	122 5 Q7YXW3	Q7YXW3 cryptosporid
43	40	59.7	122 5 Q81J44	Q81J44 plasmidium
44	40	59.7	122 15 Q7SNN1	Q7SNN1 human immun
45	40	59.7	122 15 Q7SNN0	Q7SNN0 human immun

## ALIGNMENTS

RESULT 1	PRELIMINARY	PRT	272 AA.
ID Q9TZ13	Q9TZ13		
AC Q9TZ13	01-MAY-2000 (TREMBLrel. 13, Created)		
DT 01-MAY-2000	(TREMBLrel. 13, Last sequence update)		
DT 01-OCT-2003	(TREMBLrel. 25, Last annotation update)		
DE F59H5.3 protein.			
GN F59H5.3			
OS Caenorhabditis elegans.			
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae.			
OC Rhabditidae; Pelodexinae; Caenorhabditis.			
OX NCBI_TaxID=6239;			
RN [1]			
RP SEQUENCE FROM N.A.			
RC STRAIN=Bristol N2;			
RX MEDLINE=94150718; PubMed=7906398;			
RA Wilson R., Alincough R., Anderson K., Baynes C., Berks M., Coulson A., Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Fulton L., Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L., Jones M., Kerhaw J., Kirsten J., Laister N., Latreille P., Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M., Parsons N., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R., Smaison N., Smith A., Sonhammer E., Staden R., Sulston J., Thierly-Mieg J., Thomas K., Vaubin M., Vaughan K., Waterston R., Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P., "2.2 Mb of contiguous nucleotide sequence from chromosome III of C. elegans."			
RT Nature 368:32-38(1994).			
RL [2]			
RN SEQUENCE FROM N.A.			
RC STRAIN=Bristol N2;			
RA Jones K., Graves T., Antoniou B., "The sequence of C. elegans cosmid F59H5."			
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.			
RN [3]			
RP SEQUENCE FROM N.A.			
RC STRAIN=Bristol N2;			

Waterston R.  
 RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF098991; AAC67450.1; -  
 DR PIR: T33515; T33515.  
 DR WormBep: F59H5.3; CE19911.  
 DR GO: GO:000515; F:protein binding; IEA.  
 DR InterPro: IPR000210; BRP\_P0Z.  
 DR InterPro: IPR002083; MATH.  
 DR InterPro: IPR008974; Traf\_dcm.  
 DR Pfam: PF00651; BTB; 1.  
 DR Pfam: PF00917; MATH; 1.  
 DR SMART: SMO0061; MATH; 1.  
 SQ SEQUENCE 272 AA; 31260 MW; 89421E02355F7241 CRC64;

Query Match  
 Best Local Similarity 77.1%; Score 47; DB 5; Length 272;  
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 Qy 4 RMDKSVNDY 12  
 Db 108 RMDKLINDY 116

RESULT 2  
 Q9F445 PRELIMINARY; PRT; 722 AA.  
 AC Q9F445;  
 DT 01-MAR-2001 (TREMBLrel. 16, Created)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
 DE Salt.  
 GN SALT.  
 OS Streptococcus salivarius.  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=1304;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=20P3;  
 RX MEDLINE=93362982; PubMed=8357242;  
 RA Rose K.F., Ronson C.W., Tagg J.R.;  
 RT "Isolation and characterization of the lantibiotic salivarin A and  
 its structural gene sala from Streptococcus salivarius 20P3."  
 RL Appl. Environ. Microbiol. 59:2014-2021(1993).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=20P3;  
 RX MEDLINE=21289078; PubMed=11395456;  
 RA Upton M., Tagg J.R., Mescombe P., Jenkinson H.F.;  
 RT "Intra- and interspecies signaling between Streptococcus salivarius  
 and Streptococcus pyogenes mediated by sala and salaI lantibiotic  
 peptides."  
 RL J. Bacteriol. 183:3931-3938(2001).  
 CC -1 SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.  
 DR EMBL: AY005472; AAC32538.1; -  
 DR GO: GO:0016021; C:integral to membrane; IEA.  
 DR GO: GO:0005524; F:ATP binding; IEA.  
 DR GO: GO:0004009; F:ATP-binding cassette (ABC) transporter acti. .; IEA.  
 DR GO: GO:0000166; F:nucleotide binding; IEA.  
 DR GO: GO:0008233; F:peptidase activity; IEA.  
 DR GO: GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR GO: GO:0006810; P:transport; IEA.  
 DR InterPro: IPR003593; AAA\_ATPase.  
 DR InterPro: IPR001140; ABC\_TM\_transp.  
 DR InterPro: IPR003439; ABC\_transporter.  
 DR InterPro: IPR005074; Peptidase\_C39.  
 DR Pfam: PF00664; ABC\_membrane; 1.  
 DR Pfam: PF00005; ABC\_tran; 1.  
 DR Pfam: PF03412; Peptidase\_C39; 1.  
 DR ProDom: PD000006; ABC\_transporter; 1.  
 DR SMART: SMO0382; AAA; 1  
 DR PROSITE: PS00211; ABC\_TRANSPORTER\_1; 1.  
 DR PROSITE: PS50893; ABC\_TRANSPORTER\_2; 1.

KW ATP-binding; Transport.  
 SQ SEQUENCE 722 AA; 83364 MW; 4CB2F4379EA711E6 CRC64;

Query Match  
 Best Local Similarity 70.1%; Score 47; DB 2; Length 722;  
 Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 1 LITRMDKSVNDY 12  
 Db 364 LISKMDKSPFDY 375

RESULT 3  
 Q9YKP4 PRELIMINARY; PRT; 858 AA.  
 AC Q9YKP4;  
 DT 01-MAY-1999 (TREMBLrel. 10, Created)  
 DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)  
 DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)  
 DE Env.  
 GN ENV.  
 OS Human immunodeficiency virus 1.  
 OC Viruses; Retroid viruses; Retroviridae; Lentivirus.  
 OX NCBI\_TaxID=11676;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C-96BW16B01;  
 RX MEDLINE=99214383; PubMed=10196340;  
 RA Novitsky V.A., Montano M.A., McLane M.F., Renjifo B., Vannberg F.,  
 RA Foley B.T., Ndung'u T.P., Rahman M., Makheima M.J., Marlink R.,  
 RA Essex M.;  
 RT "Molecular cloning and phylogenetic analysis of human immunodeficiency  
 RT virus type 1 subtype C: a set of 23 full-length clones from  
 RT Botswana."  
 RL J. Virol. 73:4427-4432(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C-96BW16B01;  
 RA Novitsky V.A., Montano M.A., McLane M.F., Renjifo B., Vannberg F.,  
 RA Foley B.T., Ndung'u T.P., Marlink R., Essex M.,  
 RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.

DR EMBL: AF110976; AAD17150.1; -  
 DR GO: GO:0016021; C:integral to membrane; IEA.  
 DR GO: GO:0019028; C:viral capsid; IEA.  
 DR GO: GO:0019031; C:viral envelope; IEA.  
 DR GO: GO:0005198; F:structural molecule activity; IEA.  
 DR InterPro: IPR000328; Env\_GP41.  
 DR InterPro: IPR000777; GP120.  
 DR Pfam: PF00516; GP120; 1.  
 DR Pfam: PF00517; GP41; 1.  
 KW AIDS; Coat protein; Glycoprotein; Polyprotein; Transmembrane.  
 SQ SEQUENCE 858 AA; 97264 MW; 4E069DB969338A9 CRC64;

Query Match  
 Best Local Similarity 62.7%; Score 42; DB 15; Length 858;  
 Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Qy 3 TRMDKSVNDY 12  
 Db 631 TQMDREINNY 640

RESULT 4  
 Q857E5 PRELIMINARY; PRT; 77 AA.  
 AC Q857E5;  
 DT 01-JUN-2003 (TREMBLrel. 24, Created)  
 DT 01-JUN-2003 (TREMBLrel. 24, Last sequence update)  
 DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)  
 DE Gp77.  
 GN Mycobacteriophage Bx22.  
 OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae.  
 OX NCBI\_TaxID=205870;

```

RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=2292660; PubMed=12705866;
RA Pedulla M.L., Ford M.E., Houtz J.M., Karchikyan T., Madsworth C.,
RA Lewis J.A., Jacobs-Sera D., Falbo J., Gross J., Pannunzio N.R.,
RA Brucker W., Kumar V., Kandasamy U., Keenan U., Baratarov S.,
RA Krikavsky J., Lawrence J.G., Jacobs W.R. Jr., Hendrix R.W.,
RA Hatfull G.F.;
RT "Origins of highly mosaic mycobacteriophage genomes.";
RL Cell 113:171-182(2003).
DR EMBL: AY129332; AAN0183.1;
SQ SEQUENCE 77 AA; 8809 MW; BBD295DB02C82ED6 CRC64;

Query Match 61.2%; Score 41; DB 9; Length 77;
Best Local Similarity 50.0%; Pred. No. 12;
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 3 TRMDKSVNDY 12
DB 43 SRWDEATIDY 52

RESULT 5
Q92K72 PRELIMINARY; PRT; 99 AA.
ID Q92K72;
AC Q92K72;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein.
GN SAV2532.
OS Streptomyces avermitilis.
OC Bacteria; Actinobacteria; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=33903;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=NA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
RX MEDLINE=21477403; PubMed=11572948;
RA Omura S., Ikeda H., Ishikawa U., Hanamoto A., Takahashi C.,
RA Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osone T.,
RA Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
RT "Genome sequence of an industrial microorganism Streptomyces
RT avermitilis: deducing the ability of producing secondary
RT metabolites.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=NA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
RX MEDLINE=22608306; PubMed=12692562;
RA Ikeda H., Ishikawa U., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,
RA Sakaki Y., Hattori M., Omura S.;
RT "Complete genome sequence and comparative analysis of the industrial
RT microorganism Streptomyces avermitilis.";
RL Nat. Biotechnol. 21:526-531(2003).
DR EMBL: AP005031; BAC70243.1;
KM Hypothetical protein; Complete proteome.
SQ SEQUENCE 99 AA; 10577 MW; 8B53C6DB2CB4BC CRC64;

Query Match 61.2%; Score 41; DB 16; Length 99;
Best Local Similarity 60.0%; Pred. No. 16;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 ITRMDKSVND 11
DB 59 LARWDXEAND 68

RESULT 6
Q9K693 PRELIMINARY; PRT; 479 AA.
ID Q9K693;
AC Q9K693;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)

```

```

DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE 4-hydroxyphenylacetate-3-hydroxylase.
GN BH3836.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=66665;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=C-125 / JCM 9153;
RX MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Maeno N.,
RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
RA Horikoshi K.;
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
RT halodurans and genomic sequence comparison with Bacillus subtilis.";
RL Nucleic Acids Res. 28:4317-4331(2000).
DR EMBL: AP001520; BAB07555.1;
DR PIR: D84129; D84129.
DR GO: GO:0016491; F:oxidoreductase activity; IEA.
DR InterPro: IPR004925; HpaB.
DR Pfam: PF03241; HpaB; 1.
KM Complete proteome.
SQ SEQUENCE 479 AA; 54604 MW; CFBDC437FE2E2066 CRC64;

Query Match 61.2%; Score 41; DB 16; Length 479;
Best Local Similarity 50.0%; Pred. No. 93;
Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 LITRMDKSVNDY 12
DB 59 LITCWDERTNEY 70

RESULT 7
Q92R00 PRELIMINARY; PRT; 562 AA.
ID Q92R00;
AC Q92R00;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN F4C21.24 OR A14903310.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RX Habermann K., de la Bastide M., Huang E.N., Gnoj L., Schutz K.,
RX Preston R., Calma C., Martensen R., Parnell L.D., Dedhia N.,
RX McCormie W.R.;
RT "Arabidopsis thaliana BAC F4C21 from chromosome IV near 17 cM.";
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX Spiegel L.A., Huang E.N., Nascimento L.V., de la Bastide M., Vil D.M.,
RX Shekter M., Schutz K., Matero A., Shah R., O'Shaughnessy A., Rodriguez M.,
RX Shekter M., Schutz K., See L.H., Swaby I., Habermann K., Dedhia N.N.,
RX Mewes H.W., Lemcke K., Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RX EU Arabidopsis sequencing project;
RX Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AC005275; RAD14459.1;
DR EMBL: AL161496; CAB77816.1;
DR PIR: A85042; A85042.
DR GO: GO:0016020; C:membrane; IEA.
DR GO: GO:0008233; F:peptidase activity; IEA.
DR GO: GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro: IPR000508; Peptidase_S26.

```

DR InterPro; IPR004252; Transposase\_24.  
 DR Pfam; PF03004; Transposase\_24; 1.  
 DR PROSITE; PS00761; SPASE\_1\_3; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 562 AA; 63125 MW; 206CA70981F2941A CRC64;

Query Match  
 Best Local Similarity 61.2%; Score 41; DB 10; Length 562;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 5 WDKSVND 11  
 Db 149 WDKSIND 155

RESULT 8  
 ID Q90Q08 PRELIMINARY; PRT; 865 AA.  
 AC Q90Q08;  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)  
 DE Envelope glycoprotein (Fragment).  
 DE ENV.  
 OS Human immunodeficiency virus 1.  
 OC Viruses; Retroid viruses; Retroviridae; Lentivirus.  
 OC NCBI\_TaxID=11676;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=94CY042-10e;  
 RA Gao F., Vidal N., Li Y., Trask S.A., Chen Y., Kostrikis L.G., Ho D.D.,  
 RA Kim J., Choe K., Oh M.-D., Salmunen M., Robertson D.L., Shaw G.M.,  
 RA Hahn B.H., Peeters M.,  
 RT "Evidence for two distinct sub-subtypes within the HIV-1 subtype A  
 RT radiation."  
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF286249; AAK62673.1; -  
 DR GO; GO:0016021; C:Integral to membrane; IEA.  
 DR GO; GO:0019028; C:Viral capsid; IEA.  
 DR GO; GO:0019031; C:Viral envelope; IEA.  
 DR GO; GO:0005198; F:Structural molecule activity; IEA.  
 DR InterPro; IPR000328; Env GP41.  
 DR InterPro; IPR000777; GP120.  
 DR Pfam; PF00516; GP120; 1.  
 DR Pfam; PF00517; GP41; 1.  
 KW AIDS; Coat protein; Glycoprotein; Polypeptide; Transmembrane.  
 FT NON TER 1 1  
 SQ SEQUENCE 865 AA; 97687 MW; C32EE79B7BA71F3C CRC64;

Query Match  
 Best Local Similarity 61.2%; Score 41; DB 15; Length 865;  
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 4 RMDKSVND 12  
 Db 632 QMDKESVD 640

RESULT 9  
 ID Q8T1W0 PRELIMINARY; PRT; 76 AA.  
 AC Q8T1W0;  
 DT 01-JUN-2002 (TREMBLrel. 21, Created)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Predicted protein.  
 DE MA4031.  
 OS Methanosarcina acetivorans.  
 OC Archaea; Euryarchaeota; Methanosarcinales; Methanosarcina.  
 OC NCBI\_TaxID=2214;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RC STRAIN=C2A / ATCC 35395 / DSM 2834;  
 RX MEDLINE=21929760; PubMed=11932238;  
 RA Galagan J.E., Nussbaum C., Roy A., Endrizzi M.G., Macdonald P.,  
 RA Fitzhugh W., Calvo S., Engels R., Smirnov S., Altrock D., Brown A.,  
 RA Allen N., Naylor J., Stange-Thomann N., DeLellano K., Johnson R.,  
 RA Linton L., McSwan P., McKernan K., Talamas J., Tirelli A., Ye W.,  
 RA Zimmer A., Barber R.D., Cann I., Graham D.E., Grahame D.A., Guss A.M.,  
 RA Hedderich R., Ingram-Smith C., Kuetner H.C., Krzycki J.A.,  
 RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,  
 RA Springer T.A., Umayam L.A., White O., White R.H., de Macario E.C.,  
 RA Perry J.G., Jarrell K.F., Jing H., Macario A.J.L., Paulsen I.,  
 RA Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,  
 RA Metcalf W.W., Birren B.,  
 RT "The genome of Methanosarcina acetivorans reveals extensive metabolic  
 RT and physiological diversity."  
 RL Genome Res. 12:532-542(2002).  
 DR EMBL; AB01115; AAM07379.1; -  
 KW Complete proteome.  
 SQ SEQUENCE 76 AA; 9568 MW; 8300EBA0D57985F CRC64;

Query Match  
 Best Local Similarity 59.7%; Score 40; DB 17; Length 76;  
 Matches 5; Conservative 4; Mismatches 3; Indels 0; Gaps 0;  
 QY 1 LTRMDKSVND 12  
 Db 45 LTRMEKIXD 56

RESULT 10  
 ID Q9YRS3 PRELIMINARY; PRT; 117 AA.  
 AC Q9YRS3;  
 DT 01-MAY-1999 (TREMBLrel. 10, Created)  
 DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)  
 DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)  
 DE Envelope glycoprotein (Fragment).  
 DE ENV.  
 OS Human immunodeficiency virus 1.  
 OC Viruses; Retroid viruses; Retroviridae; Lentivirus.  
 OC NCBI\_TaxID=11676;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=96USNG62;  
 RA Sullivan P.J., Do A., Ellenberger D.L., Pau C.-F., Paul S., Kalish M.,  
 RA Robbins K., Lal R., Stoeck C., Schable C.A., Wise H., Tetch C.,  
 RA Jones J., Ward J.,  
 RT "Surveillance of Central African Nationals living in the United States  
 RT reveals multiple subtypes of HIV-1 Group M and Group O."  
 RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF096343; AAD04418.1; -  
 DR GO; GO:0016021; C:Integral to membrane; IEA.  
 DR GO; GO:0019031; C:Viral envelope; IEA.  
 DR GO; GO:0005198; F:Structural molecule activity; IEA.  
 DR InterPro; IPR000328; Env\_GP41.  
 DR Pfam; PF00517; GP41; 1.  
 DR Transmembrane.  
 FT NON TER 1 1  
 FT NON TER 117 117  
 SQ SEQUENCE 117 AA; 13948 MW; 37D0A7BE32D777B CRC64;

Query Match  
 Best Local Similarity 59.7%; Score 40; DB 15; Length 117;  
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 4 RMDKSVND 12  
 Db 70 QMDKEINNY 78

RESULT 11  
 ID Q9YRU2 PRELIMINARY; PRT; 117 AA.

```

AC Q9YRU2;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Envelope glycoprotein (Fragment).
GN ENV.
OS Human immunodeficiency virus 1.
OC Viruses; Retroviridae; Retroviridae; Lentivirus.
OC NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=96USCP92;
RA Sullivan P.J., Do A., Ellenberger D.L., Pau C.-P., Paul S., Kalish M., Robbins K., Lal R., Storch C., Schable C.A., Wise H., Tetteh C., Jones J., Ward J.
RT "Surveillance of Central African Nationals living in the United States reveals multiple subtypes of HIV-1 Group M and Group O."
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF096323; AAD04398.1; -.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0019031; C:viral envelope; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR000328; Env_GP41.
DR Pfam; PF00517; GP41; 1.
KW Transmembrane.

FT NON TER 1 1
FT SEQUENCE 117 AA; 14024 MW; A6D02B387651DD39 CRC64;

Query Match
Best Local Similarity 59.7%; Score 40; DB 15; Length 117;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 4 RMDKSVNDY 12
Db 70 QMDKEINNY 78

RESULT 12
Q9ESR4 PRELIMINARY; PRT; 118 AA.
AC Q9ESR4;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Envelope glycoprotein (Fragment).
GN ENV.
OS Human immunodeficiency virus 1.
OC Viruses; Retroviridae; Retroviridae; Lentivirus.
OC NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CAM170;
RA MEDLINE=20414627; PubMed=10957729;
RA Peter F.N., Bittel M.N., Judith T.N., George A.A., Laura E.T., John N.N., Feng G., Mark R., Thomas F.M., Danuta P., Renu L.B.;
RT "Presence of Diverse Human Immunodeficiency Virus Type 1 Viral Variants in Cameroon."
RL AIDS Res. Hum. Retroviruses 16:1319-1324(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CAM170;
RA Ngenasong J.N., Gao F., Rayfield M., Folke T.M., Pieniazek D., Lal R.B.;
RT Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF252107; AAG14316.1; -.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0019031; C:viral envelope; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR000328; Env_GP41.
DR Pfam; PF00517; GP41; 1.
KW Transmembrane.

```

```

FT NON TER 1 1
FT SEQUENCE 118 AA; 14188 MW; 34BAE7C5BC91CC8 CRC64;

Query Match
Best Local Similarity 59.7%; Score 40; DB 15; Length 118;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 4 RMDKSVNDY 12
Db 66 QMDKEINNY 74

RESULT 13
Q9ESQ8 PRELIMINARY; PRT; 118 AA.
AC Q9ESQ8;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Envelope glycoprotein (Fragment).
GN ENV.
OS Human immunodeficiency virus 1.
OC Viruses; Retroviridae; Retroviridae; Lentivirus.
OC NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CAM203;
RA MEDLINE=20414627; PubMed=10957729;
RA Peter F.N., Bittel M.N., Judith T.N., George A.A., Laura E.T., John N.N., Feng G., Mark R., Thomas F.M., Danuta P., Renu L.B.;
RT "Presence of Diverse Human Immunodeficiency Virus Type 1 Viral Variants in Cameroon."
RL AIDS Res. Hum. Retroviruses 16:1319-1324(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CAM203;
RA Fonjongo P.N., Mpoudi E.N., Torimiro J.N., Alemuji G.A., Eno L.T., Ngenasong J.N., Gao F., Rayfield M., Folke T.M., Pieniazek D., Lal R.B.;
RT Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF252113; AAG14322.1; -.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0019031; C:viral envelope; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR000328; Env_GP41.
DR Pfam; PF00517; GP41; 1.
KW Transmembrane.
FT NON TER 1 1
FT SEQUENCE 118 AA; 14173 MW; 488033169BCD3CB6 CRC64;

Query Match
Best Local Similarity 59.7%; Score 40; DB 15; Length 118;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 4 RMDKSVNDY 12
Db 66 QMDKEINNY 74

RESULT 14
Q9DDO4 PRELIMINARY; PRT; 122 AA.
AC Q9DDO4;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Envelope glycoprotein (Fragment).
GN ENV.
OS Human immunodeficiency virus 1.
OC Viruses; Retroviridae; Retroviridae; Lentivirus.
OC NCBI_TaxID=11676;

```

```

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=UG;
RA Hu D.J., Baggs J., Downing R.G., Pieniazek D., Dorn J., Fridlund C.,
RT Bityahwaho B., Sempala S.D.K., Rayfield M.A., Dondero T.J., Lal R.,
RT "Similar distribution and continued predominance of HIV-1 subtypes A
and D infections in Uganda."
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF307677; AA08738.1; -.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0019031; C:viral envelope; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR000328; Env_GP41.
DR Pfam; PF00517; GP41; 1.
KW Transmembrane.
FT NON_TER 1
FT NON_TER 122
SQ SEQUENCE 122 AA; 14678 MW; 4C407CA763CC33F0 CRC64;

```

```

Query Match 59.7%; Score 40; DB 15; Length 122;
Best Local Similarity 55.6%; Pred. No. 30;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 4 RMDKSVNDY 12
DB 70 QMDKEISDY 78

```

## RESULT 15

```

Q90DM1 PRELIMINARY; PRT; 122 AA.
AC Q90DM1
DT 01-DEC-2001 (TREMBLrel. 19; Created)
DT 01-DEC-2001 (TREMBLrel. 19; Last sequence update)
DT 01-JUN-2003 (TREMBLrel. 24; Last annotation update)
DE Envelope glycoprotein (Fragment).
GN ENV.
OS Human immunodeficiency virus 1.
OC Viruses; Retroviruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=UG;
RA Hu D.J., Baggs J., Downing R.G., Pieniazek D., Dorn J., Fridlund C.,
RT Bityahwaho B., Sempala S.D.K., Rayfield M.A., Dondero T.J., Lal R.,
RT "Similar distribution and continued predominance of HIV-1 subtypes A
and D infections in Uganda."
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF307710; AA08771.1; -.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0019031; C:viral envelope; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR000328; Env_GP41.
DR Pfam; PF00517; GP41; 1.
KW Transmembrane.
FT NON_TER 1
FT NON_TER 122
SQ SEQUENCE 122 AA; 14645 MW; 67BB2CC84F41D34 CRC64;

```

```

Query Match 59.7%; Score 40; DB 15; Length 122;
Best Local Similarity 55.6%; Pred. No. 30;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 4 RMDKSVNDY 12
DB 70 QMDKEISDY 78

```

Search completed: October 5, 2004, 08:13:27  
 Job time: 11.3519 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

CM protein - protein search, using sw model

Run on: October 5, 2004, 07:04:10 ; Search time 9.12963 Seconds  
(without alignments)  
371.381 Million cell updates/sec

Title: US-09-805-290A-13  
Perfect score: 67  
Sequence: 1 LITRMDKSVNDY 12

Scoring table: BLOSUM62  
Gapop 10.0 ; Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: Geneseqp1980a:.\*  
2: Geneseqp1990a:.\*  
3: Geneseqp2000a:.\*  
4: Geneseqp2001a:.\*  
5: Geneseqp2002a:.\*  
6: Geneseqp2003a:.\*  
7: Geneseqp2003bs:.\*  
8: Geneseqp2004a:.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	67	100.0	12	4	AAE10546
2	67	100.0	129	4	AAE10553
3	45	67.2	1953	6	AAE35490
4	40	59.7	268	2	AAE22876
5	40	59.7	268	5	ABG68347
6	40	59.7	268	5	ABU57754
7	40	59.7	532	3	ADSE8141
8	40	59.7	855	3	AAE68351
9	40	59.7	1659	6	ABU00861
10	40	59.7	1659	6	ABP81332
11	40	59.7	1815	4	ABE68268
12	39	58.2	74	4	AAU43975
13	39	58.2	74	6	ABM40494
14	39	58.2	527	5	ABR08079
15	39	58.2	1191	6	ABU24124
16	39	58.2	134	2	AAE38071
17	38	56.7	134	2	AAU39214
18	38	56.7	145	4	AAU50508
19	38	56.7	145	6	ABM47027
20	38	56.7	343	4	ABE71109
21	38	56.7	665	6	ABU11554
22	38	56.7	1089	6	ABG79939
23	38	56.7	1461	6	ABP57442
24	38	56.7	2659	6	ABE68426
25	37	55.2	107	3	AAE14603

26	37	55.2	175	5	AAO14414	AAO14414 Streptococ
27	37	55.2	175	6	ABU44108	ABU44108 Protein e
28	37	55.2	294	6	ABM71201	ABM71201 Staphyloc
29	37	55.2	337	5	ABE62784	ABE62784 Protein f
30	37	55.2	360	6	ABR57597	ABR57597 HIV gp41
31	37	55.2	388	6	ABE69046	ABE69046 Phototrab
32	37	55.2	414	6	ABU35485	ABU35485 Protein e
33	37	55.2	471	6	AAO16298	AAO16298 Human PHE
34	37	55.2	529	5	ABP35688	ABP35688 Fungal ZB
35	37	55.2	609	5	ABP73457	ABP73457 Candida a
36	37	55.2	664	5	AAO26241	AAO26241 MDPF rela
37	37	55.2	749	3	AAE84627	AAE84627 Amino ac1
38	37	55.2	749	3	AAE61190	AAE61190 Human PHE
39	37	55.2	749	6	ABR58360	ABR58360 BCU0627 P
40	37	55.2	749	7	ADC77699	ADC77699 Human 956
41	37	55.2	833	2	AAE28960	AAE28960 Delta D11
42	37	55.2	833	2	AAO27056	AAO27056 Delta pro
43	37	55.2	833	4	ABM71247	ABM71247 Drosophi1
44	37	55.2	834	3	AAE79031	AAE79031 Drosophi1
45	37	55.2	846	3	ABE69353	ABE69353 HIV-1 non

## ALIGNMENTS

RESULT 1	
AAE10546	AAE10546 standard; peptide; 12 AA.
ID	AAE10546; standard; peptide; 12 AA.
AC	AAE10546;
DT	10-DEC-2001 (first entry)
DE	Llama species antibody VHH CDR3 #15.
XX	Llama antibody; camelid; anorectic; heavy chain variable domain; VHH;
KW	human dietary enzyme inhibitor; medicament; human pancreatic lipase; HPL;
KW	food; human gastric lipase; HGL; cosmetic control; body weight;
KW	complementarity determining region 3; CDR3.
OS	Llama sp.
XX	
PN	EP1134231-A1.
PD	19-SEP-2001.
XX	
PF	20-FEB-2001; 2001EP-00200703.
XX	
PR	14-MAR-2000; 2000EP-00200930.
XX	
PA	(UNITL) UNILEVER NV.
PA	(UNITL) UNILEVER PLC.
PI	Bezemer S, Van De Burg M, De Haard JWM, Tareilus E;
DR	WPI; 2001-572718/65.
XX	
PT	New antibody or its fragments for inhibiting human dietary enzymes,
PT	useful for cosmetic control of body weight of human beings, comprises
PT	heavy chain variable domain derived from immunoglobulin naturally devoid
PT	of light chains.
XX	
PS	Claim 4; Page 29; 37pp; English.
XX	
CC	The patent discloses antibodies or their fragments comprising a heavy
CC	chain variable domain (VHH) derived from an immunoglobulin naturally
CC	devoid of light chains specific for inhibiting human dietary enzymes. The
CC	antibodies of the invention are useful for the preparation of medicaments
CC	or food for inhibiting the activity of one or more human dietary enzymes
CC	especially human pancreatic lipase (HPL) or human gastric lipase (HGL)
CC	which are useful for the cosmetic control of body weight of human beings.
CC	The present peptide sequence is a complementarity determining region 3
CC	(CDR3) of llama species (camelid) antibody VHH region





XX SQ Sequence 1953 AA;  
 XX Query Match 67.2%; Score 45; DB 6; Length 1953;  
 XX Best Local Similarity 77.8%; Pred. No. 1.6e+02;  
 XX Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 OY 4 RMDKSVNDY 12  
 DB 1540 RMDKSLGDP 1548  
 RESULT 4  
 ID AAY22876 standard; protein; 268 AA.  
 AC AAY22876;  
 DT 19-AUG-1999 (first entry)  
 DE SEQ ID NO. 72 from WO9820036.  
 XX HIV; gp41 protein; constrained helical peptide; HIV infection; vaccine;  
 XX antibody; viral membrane fusion; viral infectivity;  
 XX ligand affinity purification; protein A replacement;  
 XX immunoglobulin purification; epitope mimic.  
 OS Human immunodeficiency virus.  
 PN WO9820036-A1;  
 PD 14-MAY-1998.  
 PF 05-NOV-1997; 97WO-US020069.  
 PR 06-NOV-1996; 96US-00743698.  
 PR 16-JUN-1997; 97US-00876698.  
 XX (GETH ) GENENTECH INC.  
 XX Braisted A, Judice JK, McDowell RS, Phelan JC, Starovasnik MA;  
 PI Wells JA;  
 DR WPI; 1998-286866/25.  
 XX Production of constrained helical peptide(s) by linking side chains on  
 PT termini of octapeptide - derived from human immunodeficiency virus gp41  
 PT protein, useful in vaccines for treatment and prevention of infection.  
 PS Claim 11; Page 206-207; 279pp; English.

SQ Sequence 268 AA;  
 Query Match 59.7%; Score 40; DB 2; Length 268;  
 Best Local Similarity 55.6%; Pred. No. 1.2e+02;  
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
 OY 4 RMDKSVNDY 12  
 DB 160 QMDKEINNY 168  
 RESULT 5  
 ID ABG68347 standard; protein; 268 AA.  
 AC ABG68347;  
 DT 29-AUG-2003 (revised)  
 DT 07-OCT-2002 (first entry)  
 DE Envelope protein gp41 from HIV clade A strain #5.  
 XX HIV; glycoprotein; gp41; antigen; helical conformation;  
 XX virus-induced membrane fusion; acquired immunodeficiency syndrome; AIDS;  
 XX viral envelope protein; vaccine; virucide; anti-HIV.  
 OS Human immunodeficiency virus 1; clade A.  
 PN US6271198-B1.  
 PD 07-AUG-2001.  
 PF 05-NOV-1997; 97US-00965056.  
 PR 06-NOV-1996; 96US-00743698.  
 PR 16-JUN-1997; 97US-0049287P.  
 PR 16-JUN-1997; 97US-00876698.  
 XX (GETH ) GENENTECH INC.  
 XX Braisted AC, Judice JK, McDowell RS, Phelan JC, Starovasnik MA;  
 PI Wells JA;  
 DR WPI; 2002-487624/52.  
 XX New cyclic peptides from human immune deficiency virus gp41, useful for  
 PT treatment or prevention of HIV infection, are constrained to have alpha-  
 PT helical conformation.  
 PS Disclosure; Col 227-230; 175pp; English.

XX The invention relates to cyclic peptides (A) with a constrained helical  
 CC conformation, derived from gp41 (glycoprotein 41, a viral envelope  
 CC protein) of human immunodeficiency virus (HIV). The cyclic  
 CC peptides have formulas given in the specification part of which are  
 CC derived from a consensus sequence of gp41 derived from HIV clades A, B,  
 CC C, D, E or O. The peptides are used to cause induction of a specific  
 CC immune response, resulting in antibodies that prevent virus-induced  
 CC membrane fusion. The peptides are used to treat subjects with, or at risk  
 CC of, HIV infection. The peptides are used to treat subjects with, or at risk  
 CC of, HIV infection, either as antilution/anti-infection agents or,  
 CC preferably where associated with a carrier, as an immunogen (including as  
 CC vaccine) to raise antibodies. The antibodies may be used for diagnosis or  
 CC prevention/treatment of HIV infection (i.e. acquired immunodeficiency  
 CC syndrome, AIDS), e.g. prevention of mother-to-child transmission or in  
 CC cases of health care accidents. The peptides can be based on specific HIV  
 CC strains, e.g. breakthrough isolates of HIV that have developed during  
 CC vaccine trials, so a combination of them should cover a wide range of  
 CC protection. The present sequence is gp41 protein from a particular HIV  
 CC clade used to derive a consensus sequence of gp41. (Updated on 29-AUG-  
 CC 2003 to standardise OS field)  
 SQ Sequence 268 AA;

Query Match 59.7%; Score 40; DB 5; Length 268;  
 Best Local Similarity 55.6%; Pred. No. 1.2e+02;  
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 4 RWDKSVNDY 12  
 :|||:|:  
 DB 160 QWDKEINNY 168

RESULT 6  
 ABUS7754  
 ID ABUS7754 standard; protein; 268 AA.  
 XX  
 AC ABUS7754;  
 XX  
 DT 10-APR-2003 (first entry)  
 XX  
 DE Human immunodeficiency virus (HIV) envelope protein gp41 #65.  
 XX  
 KW Human immunodeficiency virus; HIV; vaccine; helical peptide compound;  
 KW viral membrane fusion; hapten; immunogen; peptidomimetic; gp41;  
 KW envelope protein.  
 XX  
 CS Human immunodeficiency virus.  
 XX  
 PN US2002151473-A1.  
 XX  
 PD 17-OCT-2002.  
 XX  
 PE 15-MAY-2001; 2001US-00854816.  
 XX  
 PF 06-NOV-1996; 96US-00743698.  
 PR 16-JUN-1997; 97US-0049787P.  
 PR 16-JUN-1997; 97US-00876698.  
 PR 05-NOV-1997; 97US-00965056.  
 XX  
 PA (BRAI/) BRAISTED A. C.  
 PA (JUDI/) JUDICE J. K.  
 PA (MCO/) MCDOWELL R. S.  
 PA (PHEL/) PHELAN J. C.  
 PA (STAR/) STAROVASNIK M. A.  
 PA (WELL/) WELLS J. A.  
 XX  
 PI Braisted AC, Judice JK, McDowell RS, Phelan JC, Starovasnik MA;  
 PI Wells JA;  
 XX  
 DR WPI; 2003-182525/18.  
 XX  
 PT Novel constrained helical peptide compound useful for prophylactically or  
 PT therapeutically treating mammal at risk for or infected with human  
 PT immunodeficiency virus.  
 XX  
 PS Disclosure; Fig 16; 180pp; English.  
 XX  
 CC The invention describes a constrained helical peptide compound (I)  
 CC comprising a first constrained helical peptide comprising a sequence of 8  
 CC amino acids (a.as) having a first and second terminal residue both  
 CC flanking an internal sequence of 6 a.as, where the terminal residues have  
 CC a side chain that are linked to each other forming a locking group to  
 CC form a constrained helical peptide. (I) is useful for preparing  
 CC antibodies that prevent viral membrane fusion, as haptens, preferably  
 CC attached to a carrier, for use as an immunogen to raise antibodies that  
 CC have a diagnostic use, as a vaccine for treatment of patients at risk of  
 CC or infected with HIV, to create combinatorial constrained helical peptide  
 CC libraries that are useful in chemical selection systems, to isolate the  
 CC binding determinants from alpha-helical binding domains of known  
 CC proteins, for determining whether a binding determinate in an alpha-  
 CC helical binding domain of a known protein can serve as a structural model  
 CC for the design of peptidomimetics, to replace intact binding proteins or  
 CC protein binding domains in the affinity purification of ligands, to mimic  
 CC epitopes in proteins to selectively raise polyclonal or monoclonal  
 CC antibodies against such individual epitopes for isolating synthetic  
 CC antibody clones with a selected binding activity from phage display

CC combinatorial libraries, to provide conformationally stable variants of  
 CC peptides or proteins which exhibit floppy or unstable alpha-helical  
 CC secondary structure at one or more sites in unstrained form under  
 CC conditions of interest. This is the amino acid sequence of an HIV  
 CC envelope protein gp41, fragments of which are used in the creation of  
 CC locked helix peptides  
 XX

SQ Sequence 268 AA;

Query Match 59.7%; Score 40; DB 6; Length 268;  
 Best Local Similarity 55.6%; Pred. No. 1.2e+02;  
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 4 RWDKSVNDY 12  
 :|||:|:  
 DB 160 QWDKEINNY 168

RESULT 7  
 ADE58141  
 ID ADE58141 standard; protein; 532 AA.  
 XX  
 AC ADE58141;  
 XX  
 DT 29-JAN-2004 (first entry)  
 XX  
 DE Rat Protein P36365, SEQ ID NO 4012.  
 XX  
 KW Rat; pain; neuronal tissue; gene therapy; spinal segmental nerve injury;  
 KW chronic constriction injury; CCI; spared nerve injury; SNI; Chung.  
 XX  
 OS Rattus norvegicus.  
 XX  
 PN W02003016475-A2.  
 XX  
 PD 27-FEB-2003.  
 XX  
 PF 14-AUG-2002; 2002WO-US025765.  
 XX  
 PR 14-AUG-2001; 2001US-0312147P.  
 PR 01-NOV-2001; 2001US-0346382P.  
 PR 26-NOV-2001; 2001US-0333347P.  
 XX  
 PA (GENO ) GEN HOSPITAL CORP.  
 PA (FARB ) BAYER AG.  
 XX  
 PI Woolf C, D'urso D, Befort K, Costigan M;  
 XX  
 DR WPI; 2003-268312/26.  
 DR  
 DR GENBANK; P36365.  
 XX  
 PT New composition comprising two or more isolated polypeptides, useful for  
 PT preparing a medicament for treating pain in an animal.  
 XX  
 PS Claim 1; Page; 1017pp; English.  
 XX  
 CC The invention discloses a composition comprising two or more isolated rat  
 CC or human polynucleotides or a polynucleotide which represents a fragment,  
 CC derivative or allelic variation of the nucleic acid sequence. Also  
 CC claimed are a vector comprising the novel polynucleotide, a host cell  
 CC comprising the vector, a method for identifying a nucleotide sequence  
 CC which is differentially regulated in an animal subjected to pain and a  
 CC kit to perform the method, an array, a method for identifying an agent  
 CC that increases or decreases the expression of the polynucleotide sequence  
 CC that is differentially expressed in neuronal tissue of a first animal  
 CC subjected to pain, a method for identifying a compound which regulates  
 CC the expression of a polynucleotide sequence which is differentially  
 CC expressed in an animal subjected to pain, a method for identifying a  
 CC compound that regulates the activity of one or more of the  
 CC polynucleotides, a method for producing a pharmaceutical composition, a  
 CC method for identifying a compound or small molecule that regulates the  
 CC activity in an animal of one or more of the polypeptides given in the  
 CC specification, a method for identifying a compound useful in treating

CC pain and a pharmaceutical composition comprising the one or more  
CC polypeptides or their antibodies. The polynucleotide or the compound that  
CC modulates its activity is useful for preparing a medicament for treating  
CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction  
CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene  
CC therapy). The sequence presented is a rat protein (shown in Table 2 of  
CC the specification) which is differentially expressed during pain. Note:  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic form directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.

CC Sequence 532 AA;

QY Query Match 59.7%; Score 40; DB 7; Length 532;  
DB Best Local Similarity 50.0%; Pred. No. 2.6e+02;  
Matches 5; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

QY 1 LITRMDKSYN 10  
DB 488 ILTOMDRITVN 497

RESULT 8  
AAB69351  
ID AAB69351 standard; protein; 855 AA.

AC AAB69351;

DT 12-SEP-2003 (revised)  
DT 20-APR-2001 (first entry)

DE HIV-1 non-subtype B clone 94CY032-3 env protein.

XX HIV-1; human immunodeficiency virus; non-subtype B; gag; pol; env; vpu;  
KM vif; vpr; tat; rev; nef; vaccine.

OS Human immunodeficiency virus 1.

XX WO200026416-A1.

PD 11-MAY-2000.

PF 25-OCT-1999; 99WC-US024837.

PR 02-NOV-1998; 98US-00184418.

XX (UABR-) UAB RES FOUND.

PI Hahn BH, Shaw GM, Gao F;

XX WPI; 2000-365651/31.

PT Novel genomic nucleic acids of non-subtype B human immunodeficiency virus  
PT type 1 useful for detecting and treating AIDS comprises a specific  
PT nucleotide sequence.

PS Claim 41; Fig 21; 131pp; English.

XX The present invention provides the protein and coding sequences for a  
CC number of human immunodeficiency virus (HIV) type 1 non-subtype B  
CC isolates. The sequences shown include the near full-length coding  
CC sequences from each isolate, and the env, pol, vif, vpr, vpu, gag, tat,  
CC rev and nef proteins. These can be used to detect the presence of HIV-1  
CC in a sample and to produce antibodies against non-subtype B HIV-1 virus.  
CC These antibodies can be used in vaccines to prevent and treat HIV  
CC infection. (Updated on 12-SEP-2003 to standardise OS field)

XX Sequence 855 AA;

QY Query Match 59.7%; Score 40; DB 3; Length 855;  
DB Best Local Similarity 55.6%; Pred. No. 4.4e+02;  
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 4 RMDKSYNDY 12  
DB 628 QMDKEINNY 636

RESULT 9  
ABU00861

ID ABU00861 standard; protein; 1659 AA.

AC ABU00861;

DT 23-OCT-2003 (revised)  
DT 11-FEB-2003 (first entry)

DE S. pneumoniae type 4 strain protein from coding region #429.

XX Bacterial meningitis; pneumonia; sepsis; otitis media; ear infection;  
KM anti-inflammatory; antibacterial; immunostimulant; auditory; respiratory;  
KM gene therapy; vaccine.

OS Streptococcus pneumoniae; type 4 strain.

XX WO200277021-A2.

PD 03-OCT-2002.

PF 27-MAR-2002; 2002WO-IB002163.

PR 27-MAR-2001; 2001GB-00007658.

PA (CHIR-) CHIRON SPA.

XX (GENO-) INST GENOMIC RES.

PI Masignani V, Tettelin H, Fraser C;

XX WPI; 2003-040579/03.

DR N-PSDB; ABX06141.

PT New proteins and nucleic acid molecules from Streptococcus pneumoniae,  
PT useful as medicaments for treating or preventing a disease or infection  
PT due to streptococcus bacteria, such as pneumonia, sepsis, otitis media or  
PT ear infection.

PS Claim 1; SEQ ID NO 858; 56pp; English.

XX The invention relates to a protein comprising or having at least 50%  
CC identity to any of the 2469 amino acid sequences, identified in the  
CC specification (available on a computer readable format), or its fragment,  
CC expressed from 2469 of 2489 identified DNA coding regions from the  
CC Streptococcus pneumoniae type 4 strain genomic sequence appearing as  
CC AB56454. Also included are an antibody which binds one of the proteins,  
CC treating a patient by administering the protein, DNA or antibody (in a  
CC composition), a kit comprising first and second primers, which are the  
CC nucleic acid cited above or fragments between nucleotides 8-100 of a  
CC sequence not defined in the specification, for amplifying a target  
CC sequence contained within a Streptococcus nucleic acid sequence, where  
CC the first primer is substantially complementary to the target sequence  
CC and the second primer is substantially complementary to the complement of  
CC the target sequence, and where the parts of the primers having  
CC substantial complementarity define the termini of the primers having  
CC be amplified, assay comprising contacting a test compound with the  
CC protein, and determining whether the test compound binds to the protein  
CC and a Streptococcus pneumoniae bacterium, where one or more genes  
CC encoding the proteins has been rendered inactive. The proteins, nucleic  
CC acid molecules, antibody and compositions are useful as medicaments for  
CC treating or preventing a disease or infection due to streptococcus  
CC bacteria, particularly S. pneumoniae, such as pneumonia, sepsis, otitis  
CC media or ear infection. They are also useful in developing vaccines,  
CC diagnostics and antibiotics. The methods are useful for identifying  
CC immunodominant proteins. The present sequence is one of the 2469 proteins  
CC expressed by the identified coding regions from the genomic sequence.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published\_pct\_sequences. (Updated on 23-OCT-2003 to  
 CC standardise OS field)  
 XX  
 SQ Sequence 1659 AA;

Query Match 59.7%; Score 40; DB 6; Length 1659;  
 Best Local Similarity 75.0%; Pred. No. 9.3e+02;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 5 WDKSVNDY 12  
 |||: |||  
 Db 414 WDKAKNDY 421

RESULT 10  
 ABB81332  
 ID ABB81332 standard; protein; 1659 AA.  
 AC ABB81332;  
 XX  
 DT 04-MAR-2003 (first entry)

DE Streptococcus pneumoniae polypeptide SEQ ID NO 249.  
 XX  
 KW Streptococcus pneumoniae; infection; otitis media; antibacterial;  
 XX diagnosis; gene therapy.

OS Streptococcus pneumoniae.  
 XX  
 PN WO200283855-A2.  
 XX  
 PD 24-OCT-2002.

PF 12-APR-2002; 2002WO-US011524.  
 XX  
 PR 16-APR-2001; 2001US-0283946P.  
 XX  
 PR 18-APR-2001; 2001US-0284443P.

PA (AMCY ) AMERICAN CYANAMID CO.  
 XX  
 PI Zagursky RJ, Masi AM, Green BA, Chakravarti DN, Russell DP;  
 XX  
 PI Woocers JJ;

DR WPI; 2003-093010/08.  
 XX  
 DR N-PSDB; ABB242180.

PT New Streptococcus pneumoniae polynucleotides, useful for treating or  
 PT preventing S. pneumoniae infections, or non-systemic diseases, e.g.  
 PT otitis media, which are induced or exacerbated by S. pneumoniae.

XX Claim 42; Page 412-419; 1091pp; English.

XX The invention relates to isolated polynucleotides (AB272147-AB242522) of  
 CC a Streptococcus pneumoniae genomic sequence, a fragment or degenerate  
 CC variant of the polynucleotide or a nucleic acid sequence 95% identical to  
 CC one of the polynucleotides. The S. pneumoniae polynucleotides and encoded  
 CC polypeptides (ABP81299-ABP81674) are useful for treating or preventing S.  
 CC pneumoniae infections or non-systemic diseases, e.g. otitis media, which  
 CC are induced or exacerbated by S. pneumoniae. These are also useful for  
 CC detecting S. pneumoniae in a biological sample or diagnosing S.  
 CC pneumoniae infection in a subject. The polynucleotides have antibacterial  
 CC activity and are useful in gene therapy

XX SQ Sequence 1659 AA;

Query Match 59.7%; Score 40; DB 6; Length 1659;  
 Best Local Similarity 75.0%; Pred. No. 9.3e+02;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 5 WDKSVNDY 12  
 |||: |||  
 Db 414 WDKAKNDY 421

RESULT 11  
 ABB68268  
 ID ABB68268 standard; protein; 1815 AA.

AC ABB68268;  
 XX  
 DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster polypeptide SEQ ID NO 31596.

XX Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical.  
 XX  
 OS Drosophila melanogaster.

PN WO200171042-A2.  
 XX  
 PD 27-SEP-2001.

PF 23-MAR-2001; 2001WO-US009231.  
 XX  
 PR 23-MAR-2000; 2000US-0191637P.

PR 11-JUL-2000; 2000US-00614150.  
 XX  
 PA (PEKE ) PE CORP NY.

PI Venter JC, Adams M, Li PMD, Myers EW;  
 XX  
 PI N-PSDB; ABB12371.

DR WPI; 2001-656860/75.  
 XX  
 DR N-PSDB; ABB12371.

PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signaling and cell-cell  
 PT interactions.

PS Disclosure; SEQ ID NO 31596; 21pp + Sequence listing; English.

XX The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signaling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (AB116176-AB130511), expressed DNA  
 CC sequences (AB101840-AB16175) and the encoded proteins (AB857737-  
 CC ABB72072). The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from Wipo at ftp.wipo.int/pub/published\_pct\_sequences

XX SQ Sequence 1815 AA;

Query Match 59.7%; Score 40; DB 4; Length 1815;  
 Best Local Similarity 54.5%; Pred. No. 1e+03; 2; Indels 0; Gaps 0;  
 Matches 6; Conservative 3; Mismatches 2;

QY 1 LITRWKSVND 11  
 |||: |||  
 Db 1229 VITRWDSPLSD 1239

RESULT 12

AAU43975  
 ID AAU43975 standard; protein; 74 AA.

AC AAU43975;  
 XX  
 DT 27-FEB-2002 (first entry)

DE Propionibacterium acnes immunogenic protein #4871.

XX SAPHO syndrome; synovitis; acne; pustulosis; hyperostosis; osteomyelitis;  
 KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA,  
 KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;

XX dermatological; osteopathic; neuroprotectant.  
 XX Propionibacterium acnes.  
 XX WO200181581-A2.  
 XX 01-NOV-2001.  
 XX 20-APR-2001; 2001WO-US012865.  
 XX 21-APR-2000; 2000US-0199047P.  
 XX 02-JUN-2000; 2000US-0208841P.  
 XX 07-JUL-2000; 2000US-0216747P.  
 XX (CORI-) CORIXA CORP.  
 XX Skeiky YAM, Persing DH, Mitcham JL, Wang SS, Bhatia A;  
 XX L'maisonneuve J, Zhang Y, Jen S, Carter D;  
 XX WPI; 2001-616774/71.  
 XX N-PSDB; AAS59521.  
 XX Propionibacterium acnes polypeptides and nucleic acids useful for  
 XX vaccinating against and diagnosing infections, especially useful for  
 XX treating acne vulgaris.  
 XX Example 1; SEQ ID NO 5170; 1069pp; English.  
 XX Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic  
 XX polypeptides. The proteins and their associated DNA sequences are used in  
 XX the treatment, prevention and diagnosis of medical conditions caused by  
 XX P. acnes. The disorders include SAPHO syndrome (synovitis, acne,  
 XX pustulosis, hyperostosis and osteomyelitis), uveitis and endophthalmitis.  
 XX P. acnes is also involved in infections of bone, joints and the central  
 XX nervous system, however it is particularly involved in the inflammatory  
 XX lesions associated with acne vulgaris. A method for detecting the  
 XX presence or absence of P. acnes in a patient comprises contacting a  
 XX sample with a binding agent that binds to the proteins of the invention  
 XX and determining the amount of bound protein in the sample. The  
 XX polypeptides may be used as antigens in the production of antibodies  
 XX specific for P. acnes proteins. These antibodies can be used to  
 XX downregulate expression and activity of P. acnes polypeptides and  
 XX therefore treat P. acnes infections. The antibodies may also be used as  
 XX diagnostic agents for determining P. acnes presence, for example, by  
 XX enzyme linked immunosorbent assay (ELISA). Note: The sequence data for  
 XX this patent did not form part of the printed specification, but was  
 XX obtained in electronic format directly from WIPO at  
 XX ftp.wipo.int/pub/published\_pct\_sequences

SO Sequence 74 AA;  
 QY Query Match 58.2%; Score 39; DB 4; Length 74;  
 QY Best Local Similarity 66.7%; Pred. No. 42;  
 QY Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 QY 2 ITRMDKSVN 10  
 QY :|||:||||  
 QY 60 VTRMNSVN 68

RESULT 13  
 ABM40494  
 ID ABM40494 standard; protein; 74 AA.  
 AC ABM40494;  
 XX 20-OCT-2003 (first entry)  
 XX Propionibacterium acnes predicted ORF-encoded polypeptide #5170.  
 XX Acne vulgaris; antisepidermoic; dermatological; antibacterial;  
 XX immunostimulant; immune response; vaccine.  
 XX

OS Propionibacterium acnes.  
 XX WO2003033515-A1.  
 XX 24-APR-2003.  
 XX 11-OCT-2002; 2002WO-US032727.  
 XX 15-OCT-2001; 2001US-00978825.  
 XX (CORI-) CORIXA CORP.  
 XX Mitcham JL, Skeiky YAM, Persing DH, Bhatia A, Maisonneuve JL;  
 XX Zhang Y, Wang S, Jen S, Lodes MJ, Benson DR, Jones R, Carter D;  
 XX Barth B, Valilve-Douglais J;  
 XX WPI; 2003-381789/36.  
 XX N-PSDB; ACF64450.  
 XX New Propionibacterium acnes polypeptides and polynucleotides encoding the  
 XX polypeptide, useful for diagnosing, preventing or treating acne vulgaris,  
 XX or for stimulating an immune response specific for a P. acnes protein.  
 XX Example 1; SEQ ID NO 5170; 1481pp; English.  
 XX The invention relates to an isolated polynucleotide (ACF64435-ACF64733)  
 XX encoding a Propionibacterium acnes protein. The invention also relates to  
 XX polypeptides encoded by the polynucleotides (ABM35624-ABM64536) and to  
 XX immunogenic fragments of P. acnes polypeptides. The invention  
 XX additionally encompasses expression vectors and host cells comprising a  
 XX polynucleotide of the invention; antibodies against polypeptides of the  
 XX invention; fusion proteins comprising a polypeptide of the invention; a  
 XX method for stimulating an immune response specific for a P. acnes  
 XX polypeptide and an isolated T cell population comprising T cells prepared  
 XX via this method; a vaccine composition (comprising P. acnes polypeptides,  
 XX polynucleotides, antibodies, fusion proteins, T cell populations, or  
 XX antigen-presenting cells that express the polypeptide); a method and kit  
 XX for detecting or determining the presence or absence of P. acnes in a  
 XX patient; and a method for inhibiting the development of P. acnes in a  
 XX patient. The P. acnes polypeptides, polynucleotides, antibodies, fusion  
 XX proteins, T cell populations or antigen-presenting cells that express the  
 XX polypeptides are useful for diagnosing, preventing or treating acne  
 XX vulgaris, or for stimulating an immune response specific for a P. acnes  
 XX protein. The polynucleotides can also be used as probes or primers for  
 XX nucleic acid hybridisation. The vaccine composition is useful for the  
 XX stimulation of an immune response against P. acnes, or for treating acne,  
 XX and the kit is useful for performing a diagnostic assay. The present  
 XX sequence represents a polypeptide predicted to be encoded by an ORF (open  
 XX reading frame) contained within the P. acnes polynucleotides of the  
 XX invention. Note: The sequence data for this patent did not form part of  
 XX the printed specification, but was obtained in electronic format directly  
 XX from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

SO Sequence 74 AA;  
 QY Query Match 58.2%; Score 39; DB 6; Length 74;  
 QY Best Local Similarity 66.7%; Pred. No. 42;  
 QY Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 QY 2 ITRMDKSVN 10  
 QY :|||:||||  
 QY 60 VTRMNSVN 68

RESULT 14  
 ABB08079  
 ID ABB08079 standard; protein; 527 AA.  
 AC ABB08079;  
 XX 10-SEP-2002 (first entry)  
 XX Maize cytochrome P450, CYP72A1.  
 XX

```

XX Maize, cytochrome P450, CYP72A1, CYP92A1; plant; gene therapy; antidote.
XX
XX Zea mays.
XX
OS
XX
XX Key Location/Qualifiers
XX Peptide 200.527
XX /note="this fragment is not provided in the protein
XX sequence indicated in the sequence listing"
XX
XX US6380465-B1.
XX
XX 30-APR-2002.
XX
XX 11-JUL-1999; 99US-00351229.
XX
XX 12-JUL-1998; 98US-0092396P.
XX
XX (KENT) UNIV KENTUCKY RES FOUND.
XX
XX Barrett M:
XX
XX WPI; 2002-470227/50.
XX N-PSDB; ABL60758.
XX
XX Nucleic acid molecules encoding a Zea mays cytochrome P450 CYP72A1,
XX useful for the production of plants resistant to heterocyclic,
XX sulfonylurea and substituted urea herbicides and organophosphate
XX insecticides e.g. Classic and Pursult.
XX
XX Claim 13; Fig 1; 21pp; English.
XX
XX The invention relates to isolated nucleic acid molecules encoding Zea
XX mays (maize) cytochrome P450 designated CYP72A1 and CYP92A1. The
XX polypeptides can be expressed by standard recombinant methodology. The
XX nucleic acids may be used in genetic engineering protocols to transform
XX plants and other eukaryotes e.g. yeast, maize (especially), soybean,
XX beet, tobacco, wheat, barley, poppy, rape, sunflower, alfalfa, sorghum,
XX rose, carnation, gerbera, carrot, tomato, lettuce, chicory, pepper, melon
XX and cabbage. They may be used in this way to confer protection against
XX heterocyclic herbicides, sulfonylurea herbicides, substituted urea
XX herbicides and/or organophosphate insecticides. The present sequence
XX represents the Z. mays CYP72A1
XX
XX Sequence 527 AA:
XX
XX Query Match 58.2%; Score 39; DB 5; Length 527;
XX Best Local Similarity 60.0%; Pred. No. 3.8e+02;
XX Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
XX
XX QY 1 LITRMDKSVN 10
XX :|||||:
XX :|||:
XX 189 MTRMDNSMS 198
XX
XX
XX RESULT 15
XX ABU24124
XX ID ABU24124 standard; protein; 1191 AA.
XX
XX AC ABU24124;
XX
XX DT 19-JUN-2003 (first entry)
XX
XX DE Protein encoded by Prokaryotic essential gene #9651.
XX
XX KW Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX
XX OS Clostridium acetobutylicum.
XX
XX WO200277183-A2.
XX
XX PN
XX PD 03-OCT-2002.
XX

```

```

PF 21-MAR-2002; 2002WO-US009107.
XX
XX 21-MAR-2001; 2001US-00815242.
XX 06-SEP-2001; 2001US-00948993.
XX 25-OCT-2001; 2001US-0342923P.
XX 08-FEB-2002; 2002US-00072851.
XX 06-MAR-2002; 2002US-0362699P.
XX
XX (ELIT-) ELITRA PHARM INC.
XX
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
XX Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX
XX WPI; 2003-029926/02.
XX N-PSDB; ACA27994.
XX
XX New antisense nucleic acids, useful for identifying proteins or screening
XX for homologous nucleic acids required for cellular proliferation to
XX isolate candidate molecules for rational drug discovery programs.
XX
XX Claim 25; SEQ ID NO 52048; 1766pp; English.
XX
XX The invention relates to an isolated nucleic acid comprising any one of
XX the 6213 antisense sequences given in the specification where expression
XX of the nucleic acid inhibits proliferation of a cell. Also included are:
XX (1) a vector comprising a promoter operably linked to the nucleic acid
XX encoding a polypeptide whose expression is inhibited by the antisense
XX nucleic acid; (2) a host cell containing the vector; (3) an isolated
XX polypeptide or its fragment whose expression is inhibited by the
XX antisense nucleic acid; (4) an antibody capable of specifically binding
XX the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
XX proliferation or the activity of a gene in an operon required for
XX proliferation; (7) identifying a gene in an operon required for
XX the gene product or that has an activity against a biological pathway
XX required for proliferation, or that inhibits cellular proliferation; (8)
XX identifying a gene required for cellular proliferation or the biological
XX pathway in which a proliferation-regulated gene or its gene product lies
XX or a gene on which the test compound that inhibits proliferation of an
XX organism acts; (9) manufacturing an antibiotic; (10) profiling a
XX compound's activity; (11) a culture comprising strains in which the gene
XX product is overexpressed or underexpressed; (12) determining the extent
XX to which each of the strains is present in a culture or collection of
XX strains; or (13) identifying the target of a compound that inhibits the
XX proliferation of an organism. The antisense nucleic acids are useful for
XX identifying proteins or screening for homologous nucleic acids required
XX for cellular proliferation to isolate candidate molecules for rational
XX drug discovery programs, or for screening homologous nucleic acids
XX required for proliferation in cells other than S. aureus, S. typhimurium,
XX K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of
XX the target prokaryotic essential genes. Note: The sequence data for this
XX patent did not form part of the printed specification, but was obtained
XX in electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pcr_sequences
XX
XX Sequence 1191 AA:
XX
XX Query Match 58.2%; Score 39; DB 6; Length 1191;
XX Best Local Similarity 50.0%; Pred. No. 9.4e+02;
XX Matches 6; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
XX
XX QY 1 LITRMDKSVNDY 12
XX :|||:
XX :|||:
XX 373 LITRMDKSVNDY 384
XX
XX

```

Search completed: October 5, 2004, 07:59:52  
Job time : 13.1296 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2004, 07:36:16 ; Search time 1.94444 Seconds  
(without alignments)  
593.639 Million cell updates/sec

Title: US-09-805-290A-14  
Perfect score: 72  
Sequence: 1 RRSNYDRSGMDY 12

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :  
1: PIR 78:\*  
2: PIR1:\*  
3: PIR2:\*  
4: PIR3:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	42	58.3	128	2 PH0952	Ig heavy chain V r
2	42	58.3	457	2 S55091	probable membrane
3	41	56.9	152	2 A87679	conserved hypotet
4	41	56.9	232	2 P00538	probable phosphogl
5	41	56.9	361	2 S30313	hypothetical prote
6	41	56.9	415	2 T00140	capsid protein - S
7	41	56.9	560	2 D90571	conserved hypotet
8	41	56.9	602	1 S46676	probable phosphoe
9	40.5	56.2	890	2 F75289	alanyl-tRNA synthe
10	40	55.6	166	2 D83230	hypothetical prote
11	40	55.6	332	2 T02789	probable membrane
12	40	55.6	341	2 T14763	pectate lyase-like
13	40	55.6	355	2 T14117	NADH2 dehydrogenas
14	40	55.6	355	2 T14120	NADH2 dehydrogenas
15	40	55.6	355	2 T14125	NADH2 dehydrogenas
16	40	55.6	361	2 S30325	hypothetical prote
17	40	55.6	475	2 G90005	hypothetical prote
18	40	55.6	481	2 D70559	probable ribosoma
19	40	55.6	481	2 H87081	30S ribosomal prot
20	40	55.6	482	2 S77660	ribosomal protein
21	40	55.6	502	2 T35743	ribosomal protein
22	40	55.6	796	2 E87636	TenB-dependent rec
23	40	55.6	838	2 T45699	hypothetical prote
24	39.5	54.9	1282	2 J04393	microbial collagen
25	39	54.2	129	2 A52132	hypothetical prote
26	39	54.2	277	2 G70809	probable thiosulfa
27	39	54.2	277	2 A87184	thiosulfate sulfur
28	39	54.2	312	2 JN0596	fibriogen-related
29	39	54.2	369	2 I49280	interleukin-2 rece

30	39	54.2	450	2 S57837	lymphoid-specific
31	39	54.2	510	2 A11381	phosphoglycerate m
32	39	54.2	511	2 D69675	phosphoglycerate m
33	39	54.2	511	2 T46865	phosphoglycerate m
34	39	54.2	554	1 AJECN	asparagine synthas
35	39	54.2	627	1 S40048	1,4-alpha-glucan b
36	39	54.2	774	2 A39832	scabrous locus (sc
37	38	52.8	209	2 S12247	NADH dehydrogenase
38	38	52.8	211	2 S21864	probable cysteine
39	38	52.8	225	2 F64454	hypothetical prote
40	38	52.8	245	2 JQ0337	allergen Der p 1 -
41	38	52.8	275	2 S51641	MYOD protein - rat
42	38	52.8	319	2 A61500	allergen Der f 1 p
43	38	52.8	348	2 T03530	cobw protein - Rho
44	38	52.8	373	2 A55718	interleukin-2 rece
45	38	52.8	481	2 C71338	probable UDP-N-ace

#### ALIGNMENTS

RESULT 1  
PH0952  
Ig heavy chain V region (G6+ CUI-SMI) - human (fragment)  
C/Species: Homo sapiens (man)  
C/Date: 17-Apr-1993 #sequence\_revision 17-Apr-1993 #text\_change 16-Aug-1996  
C/Accession: PH0952  
R/Martin, T., Duffly, S.F., Carson, D.A., Kipps, T.J.  
U. Exp. Med. 175, 983-991, 1992  
A/Title: Evidence for somatic selection of natural autoantibodies.  
A/Reference number: PH0952; MUID:92202880; PMID:1552291  
A/Accession: PH0952  
A/Status: nucleic acid sequence not shown  
A/Molecule type: DNA  
A/Residues: 1-128 <MAR>  
C/Superfamily: immunoglobulin V region; immunoglobulin homology  
C/Keywords: heterodimer; immunoglobulin  
C/KeyWords: heterodimer; immunoglobulin  
F/1-30/Region: framework 1  
F/15-98/Domain: immunoglobulin homology <IMM>  
F/31-35/Region: complementarity-determining 1  
F/36-50/Region: complementarity-determining 2  
F/51-67/Region: complementarity-determining 3  
F/68-98/Region: framework 3  
F/99-116/Region: complementarity-determining 3

Query Match 58.3%; Score 42; DB 2; Length 128;  
Best local similarity 58.3%; Pred. No. 8.6;  
Matches 7; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 RRSNYDRSGMDY 12  
DB 98 RGNNDYIMGSY 109

RESULT 2  
S55091  
probable membrane protein YMR209c - yeast (Saccharomyces cerevisiae)  
N:Alternate names: hypothetical protein YMR261.03c  
C/Species: Saccharomyces cerevisiae  
C/Date: 08-Jul-1995 #sequence\_revision 01-Sep-1995 #text\_change 19-Apr-2002  
C/Accession: S55091  
R/Dedman, K.; Brown, D.; Bowman, S.  
submitted to the EMBL Data Library, June 1995  
A/Reference number: S55089  
A/Accession: S55091  
A/Molecule type: DNA  
A/Residues: 1-457 <DED>  
A/Cross-references: EMBL:Z49809; NID:9854459; PID:9854461; GSPDB:GN00013; MIPS:YMR209c  
A/Experimental source: strain AB972  
C/Genetics:  
A/Gene: MIPS:YMR209c  
A/Cross-references: SGD:S0004822  
A/Map position: 13R

C/Keywords: transmembrane protein  
F/5-21/Domain: transmembrane #status predicted <TMM>

Query Match 58.3%; Score 42; DB 2; Length 457;  
Best Local Similarity 63.6%; Pred. No. 30;  
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 RSNVDRSGDY 12  
DB 330 RSNLDPLWGDF 340

RESULT 3  
A87679

Conserved hypothetical protein CC3467 [imported] - Caulobacter crescentus  
C/Species: Caulobacter crescentus  
C/Date: 20-Apr-2001 #sequence revision 20-Apr-2001 #text\_change 20-Apr-2001

C/Accession: A87679  
R/Nielsen, W.C.; Feldblym, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.  
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon  
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.  
Proc. Natl. Acad. Sci. U.S.A. 98, 4138-4141, 2001  
A/Title: Complete Genome Sequence of Caulobacter crescentus.  
A/Reference number: A87249; MUID:21173698; PMID:11259647  
A/Accession: A87679  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-152 <SNO>

A/Cross-references: GB:A8005673; NID:G13425187; PIDN:AAK5429.1; GSPDB:GNO0148  
A/Genetics: CC3467

Query Match 56.9%; Score 41; DB 2; Length 152;  
Best Local Similarity 65.7%; Pred. No. 15;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 6 DRSGDY 12  
DB 71 DRKMGDY 77

RESULT 4  
P00538

Probable phosphoglycerate mutase (EC 5.4.2.1), 2,3-diphosphoglycerate-independent - Bac  
C/Species: Bacillus megaterium  
C/Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 27-Oct-2003  
C/Accession: P00538  
R/Schlaepfer, B.S.; Zuber, H.  
Gene 123, 53-62, 1992  
A/Title: Cloning and sequencing of the genes encoding glyceraldehyde-3-phosphate dehydro  
comparision with corresponding sequences from thermophilic Bacillus stearothermophilus.  
A/Reference number: QJ1952; MUID:93083995; PMID:1452037  
A/Accession: P00538  
A/Molecule type: DNA  
A/Residues: 1-232 <SCH>

A/Cross-references: GB:W87647; NID:G143315; PIDN:AAA73205.1; PID:G143320  
A/Experimental source: strain DSM319  
A/Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma pulmk  
C/Keywords: intramolecular transferase; isomerase

Query Match 56.9%; Score 41; DB 2; Length 232;  
Best Local Similarity 41.7%; Pred. No. 23;  
Matches 5; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 RSNVDRSGDY 12  
DB 29 KKNFDFRMY 40

RESULT 5  
S30313

hypothetical protein 8 - Methanobacterium thermoformicum plasmid pFV1  
C/Species: Methanobacterium thermoformicum

C/Date: 19-Mar-1997 #sequence\_revision 25-Apr-1997 #text\_change 22-Oct-1999  
C/Accession: S30313  
R/Noelling, J.; van Eeden, F.J.M.; Eggen, R.I.L.; de Vos, W.M.  
Nucleic Acids Res. 20, 6501-6507, 1992

A/Title: Modular organization of related Archaeal plasmids encoding different restriction  
A/Reference number: S30302; MUID:93126090; PMID:1335177

A/Accession: S30313  
A/Status: preliminary; nucleic acid sequence not shown; translation not shown  
A/Molecule type: DNA  
A/Residues: 1-361 <NOE>  
A/Cross-references: EMBL:X68366; NID:944632; PIDN:CAA48425.1; PID:944644  
A/Note: the nucleotide sequence was submitted to the EMBL Data Library, September 1992  
C/Genetics: A/genome: plasmid pFV1

Query Match 56.9%; Score 41; DB 2; Length 361;  
Best Local Similarity 58.3%; Pred. No. 35;  
Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 RSNVDRSGDY 12  
DB 281 RSDSLPRSGPY 292

RESULT 6  
T00140

capsid protein - Staphylococcus aureus phage phi PVL

C/Species: Staphylococcus aureus phage phi PVL  
C/Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 11-May-2000  
C/Accession: T00140  
R/Kaneko, J.; Kimura, T.; Kawakami, Y.; Tomita, T.; Kamio, Y.  
Biosci. Biotechnol. Biochem. 61, 1960-1962, 1997

A/Title: Barton-Valentine leukocidin genes in a phage-like particle isolated from mitomyx  
A/Reference number: Z14119; MUID:98067870; PMID:9404084  
A/Accession: T00140  
A/Status: translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 1-415 <KAN>

A/Cross-references: EMBL:AB009866; NID:d1204727; PIDN:BAJ1880.1; PID:d1032841

Query Match 56.9%; Score 41; DB 2; Length 415;  
Best Local Similarity 63.6%; Pred. No. 40;  
Matches 7; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 RSNVDRSGDY 12  
DB 363 RSOYASMTDY 373

RESULT 7  
D90571

Conserved hypothetical protein MYPV\_4760 [imported] - Mycoplasma pulmonis (strain VAB CT)  
C/Species: Mycoplasma pulmonis  
C/Date: 24-May-2001 #sequence\_revision 24-May-2001 #text\_change 03-Aug-2001

C/Accession: D90571  
R/Chandaud, I.; Hellig, R.; Ferris, S.; Barde, V.; Samson, D.; Galisson, F.; Moszer, I.;  
Nucleic Acids Res. 29, 2145-2153, 2001  
A/Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma pulmk  
A/Reference number: A99512; MUID:21267165; PMID:11353084

A/Accession: D90571  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-560 <KUR>

A/Cross-references: GB:AL445566; PID:G14089890; PIDN:CAC13649.1; GSPDB:GNO0153  
A/Experimental source: strain VAB CTIP  
C/Genetics: A/gene: MYPV\_4760  
A/genetic code: SGC3

Query Match 56.9%; Score 41; DB 2; Length 560;  
Best Local Similarity 50.0%; Pred. No. 54;  
Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;



QY 1 RRSNYDRSGWDY 12  
:|||||:|:  
Db 177 KRNSYRNYSFY 188

RESULT 8  
S46676  
probable phosphoesterase (EC 3.1.-.-) YHR202w [similarity] - yeast (Saccharomyces cerevisiae)  
N:Alternate names: hypothetical protein H9998.1  
C:Species: Saccharomyces cerevisiae  
C>Date: 28-Oct-1994 #sequence\_revision 28-Oct-1994 #text\_change 19-Apr-2002  
C:Accession: S46676  
R:Macri, C.  
Submitted to the EMBL Data Library, February 1994  
A:Description: The sequence of S. cerevisiae cosmid 9998.  
A:Reference number: S46674  
A:Accession: S46676  
A:Molecule type: DNA  
A:Residues: 1-602 <MAC>  
A:Cross-references: EMBL:U00030, NID:g458927, PID:g458928, GSPDB:GN00008, MIPS:YHR202w  
C:Comment: This sequence has motifs characteristic of a variety of phosphoesterases.  
C:Genetics:  
A:Gene: MIPS:YHR202w  
A:Cross-references: SGD:S0001245  
A:Map position: 8R  
C:Superfamily: Saccharomycetes cerevisiae probable phosphoesterase YHR202w, phosphoesterase  
C:Keywords: hydrolase  
F:45-133/Domain: phosphoesterase core homology <PFC>

Query Match 56.9%; Score 41; DB 1; Length 602;  
Best Local Similarity 45.5%; Pred. No. 57;  
Matches 5; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 2 RSNYDRSGWDY 12  
:|||||:|:  
Db 62 QNDYDADMGDF 72

RESULT 9  
F75289  
alanyl-tRNA synthetase - Deinococcus radiodurans (strain R1)  
C:Species: Deinococcus radiodurans  
C>Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 17-Mar-2003  
C:Accession: F75289  
R:White, O.; Eichen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;  
S.; Smith, H.O.; Venter, U.C.; Fraser, C.M.  
Science 286, 1571-1577, 1999  
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.  
A:Reference number: A75250; MUID:20036896; PMID:10567265  
A:Accession: F75289  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-890 <WHI>  
A:Cross-references: GB:AE002067; GB:AE000513; NID:g6460110; PIDN:AAFI1848.1; PID:g646011  
C:Genetics:  
A:Gene: DR2300  
A:Experimental source: strain R1  
C:Superfamily: alanyl-tRNA ligase

Query Match 56.2%; Score 40.5; DB 2; Length 890;  
Best Local Similarity 61.5%; Pred. No. 1e+02;  
Matches 8; Conservative 1; Mismatches 3; Indels 1; Gaps 1;

QY 1 RRSNY-DRSGWDY 12  
:|||||:|:  
Db 188 RGENYDDTWGDY 200

RESULT 10  
D83230  
hypothetical protein PA318 [imported] - Pseudomonas aeruginosa (strain PA01)

C:Species: Pseudomonas aeruginosa  
C>Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
C:Accession: D83230  
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Bri-  
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lapidis, K.; Lim,  
.; Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho-  
A:Reference number: A82550; MUID:20437337; PMID:10984043  
A:Accession: D83230  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-166 <STO>  
A:Cross-references: GB:AE004754; GB:AE004091; NID:g949446; PIDN:AAG06706.1; GSPDB:GN0013  
A:Experimental source: strain PA01  
C:Genetics:  
A:Gene: PA3318

Query Match 55.6%; Score 40; DB 2; Length 166;  
Best Local Similarity 54.5%; Pred. No. 24;  
Matches 6; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 RSNYDRSGWDY 12  
:|||||:|:  
Db 81 RGNYDAMRQY 91

RESULT 11  
T02789  
probable membrane protein I549.1 [imported] - Leishmania major (strain Friedlin)  
C:Species: Leishmania major  
C>Date: 24-Mar-1999 #sequence\_revision 24-Mar-1999 #text\_change 19-May-2000  
C:Accession: A81455; T02789  
R:Myler, P.J.; Audleman, L.; deVos, T.; Hixson, G.; Kiser, P.; Lemley, C.; Magness, C.; F  
Proc. Natl. Acad. Sci. U.S.A. 96, 2902-2906, 1999  
A:Title: Leishmania major Friedlin chromosome 1 has an unusual distribution of protein-c-  
A:Reference number: A81455; MUID:99178987; PMID:10077609  
A:Accession: A81455  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-132 <PLY>  
A:Cross-references: GB:AE001274; NID:g3264850; PIDN:AAC24613.1; PID:g2978450; GSPDB:GN001  
C:Genetics:  
A:Gene: L549.1  
A:Map position: 1  
C:Superfamily: Leishmania major probable membrane protein I549.1  
C:Keywords: transmembrane protein

Query Match 55.6%; Score 40; DB 2; Length 332;  
Best Local Similarity 66.7%; Pred. No. 47;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 SNYDRSGWD 11  
:|||||:|:  
Db 146 TNYPRSWGN 154

RESULT 12  
T47653  
pectate lyase-like protein - Arabidopsis thaliana  
N:Alternate names: protein T26112.20  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C>Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 20-Apr-2000  
C:Accession: T47653  
R:Montfort, A.; Casacuberta, E.; Puigdomenech, P.; Mewes, H.W.; Lemcke, K.; Mayer, K.F.X.;  
submitted to the Protein Sequence Database, February 2000  
A:Reference number: Z24471  
A:Accession: T47653  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-341 <MON>  
A:Cross-references: EMBL:AL132954

A:Experimental source: cultivar Columbia, BAC clone T26112  
 C:Genetics:  
 A:Map position: 3  
 A:Introns: 43/1; 170/2; 240/3; 263/3  
 A>Note: T26112.20

Query Match 55.6%; Score 40; DB 2; Length 341;  
 Best Local Similarity 60.0%; Pred. No. 48;  
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 3 NYDRSGMDY 12  
 Db 222 NYDRSGMDY 231

## RESULT 13

T14117  
 NADH2 dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 5 - Ohomopterus mayasanus mitochondrion  
 C:Species: mitochondrion Ohomopterus mayasanus  
 A:Variety: isolate I-7  
 C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 03-Jun-2002  
 C:Accession: T14117  
 R:Su, Z.H.; Tomioka, O.; Ohama, T.; Kajiwara, E.; Ishikawa, R.; Okada, T.S.; Nakamura, J. Mol. Evol. 43, 662-671, 1996  
 A:Title: Parallel evolution in radiation of Ohomopterus ground beetles inferred from mit  
 A:Reference number: Z17791; MUID:97149035; PMID:8995063  
 A:Accession: T14117  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-355 <SUZ>  
 A:Cross-references: EMBL:D50729; NID:G1001046; PIDN:BAA09360.1; PID:G1434888  
 A:Experimental source: ssp. takiharensis; isolate I-7; adult; thorax muscle  
 C:Genetics:  
 A:Genome: mitochondrion  
 C:Superfamily: NADH dehydrogenase (ubiquinone) chain 5  
 C:Keywords: membrane-associated complex; mitochondrion; NAD; oxidative phosphorylation;

Query Match 55.6%; Score 40; DB 2; Length 355;  
 Best Local Similarity 55.6%; Pred. No. 50;  
 Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 4 NYDRSGMDY 12  
 Db 302 NYDRSGMDY 310

## RESULT 14

T14120  
 NADH2 dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 5 - Ohomopterus shigaraki mitochondrion  
 C:Species: mitochondrion Ohomopterus shigaraki  
 C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 03-Jun-2002  
 C:Accession: T14120  
 R:Su, Z.H.; Tomioka, O.; Ohama, T.; Kajiwara, E.; Ishikawa, R.; Okada, T.S.; Nakamura, J. Mol. Evol. 43, 662-671, 1996  
 A:Title: Parallel evolution in radiation of Ohomopterus ground beetles inferred from mit  
 A:Reference number: Z17791; MUID:97149035; PMID:8995063  
 A:Accession: T14120  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-355 <SUZ>  
 A:Cross-references: EMBL:D50728; NID:G1001020; PIDN:BAA09359.1; PID:G1434887  
 A:Experimental source: adult; thorax muscle  
 C:Genetics:  
 A:Genome: mitochondrion  
 C:Superfamily: NADH dehydrogenase (ubiquinone) chain 5  
 C:Keywords: membrane-associated complex; mitochondrion; NAD; oxidative phosphorylation;

Query Match 55.6%; Score 40; DB 2; Length 355;  
 Best Local Similarity 55.6%; Pred. No. 50;  
 Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 4 NYDRSGMDY 12  
 Db 302 NYDRSGMDY 310

Db 302 NYDRSGMDY 310

## RESULT 15

T14125  
 NADH2 dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 5 - Ohomopterus yaconinus mitochondrion  
 C:Species: mitochondrion Ohomopterus yaconinus  
 A:Variety: isolate Mle (Y-3)  
 C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 03-Jun-2002  
 C:Accession: T14125  
 R:Su, Z.H.; Tomioka, O.; Ohama, T.; Kajiwara, E.; Ishikawa, R.; Okada, T.S.; Nakamura, J. Mol. Evol. 43, 662-671, 1996  
 A:Title: Parallel evolution in radiation of Ohomopterus ground beetles inferred from mitc  
 A:Reference number: Z17791; MUID:97149035; PMID:8995063  
 A:Accession: T14125  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-355 <SUZ>  
 A:Cross-references: EMBL:D50731; PIDN:BAA09362.1  
 A:Experimental source: ssp. cupidicornis; isolate Mle (Y-3); adult; thorax muscle  
 C:Genetics:  
 A:Genome: mitochondrion  
 C:Superfamily: NADH dehydrogenase (ubiquinone) chain 5  
 C:Keywords: membrane-associated complex; mitochondrion; NAD; oxidative phosphorylation;

Query Match 55.6%; Score 40; DB 2; Length 355;  
 Best Local Similarity 55.6%; Pred. No. 50;  
 Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 4 NYDRSGMDY 12  
 Db 302 NYDRSGMDY 310

Search completed: October 5, 2004, 08:17:05  
 Job time: 4.9444 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2004, 07:29:35 ; Search time 1.07407 Seconds  
(without alignments)  
581.749 Million cell updates/sec

Title: US-09-805-290A-14  
Perfect score: 72  
Sequence: 1 RSNYDRSGWGY 12

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues  
Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	42	58.3	457 1 YMS9_YEAST	Q03648 saccharomyc
2	41	56.9	361 1 YPV8_METTF	P29584 methanobact
3	41	56.9	511 1 GPM1_BACME	P35167 bacillus me
4	41	56.9	602 1 YH02_YEAST	P38887 saccharomyc
5	40.5	56.6	880 1 SYA_DEIRA	O98287 delnoccocus
6	40.5	55.6	361 1 YP28_METTF	P29585 methanobact
7	40	55.6	481 1 RSI_MYCLE	P46536 mycobacteri
8	40	55.6	481 1 RSI_MYCLE	O06147 mycobacteri
9	39	54.2	277 1 THT1_MYCOB	P59899 mycobacteri
10	39	54.2	277 1 THT1_MYCOB	O50036 mycobacteri
11	39	54.2	277 1 THT1_MYCTU	O05793 mycobacteri
12	39	54.2	281 1 THT1_SACER	P16385 saccharopol
13	39	54.2	312 1 FGL1_HUMAN	Q08830 homo sapien
14	39	54.2	369 1 CYRG_MOUSE	P34502 mus musculu
15	39	54.2	450 1 IRF4_MOUSE	O64287 mus musculu
16	39	54.2	510 1 GPM1_BACST	O9x519 bacillus st
17	39	54.2	510 1 GPM1_BACST	P39773 bacillus st
18	39	54.2	510 1 GPM1_TISMO	O8y414 listeria mo
19	39	54.2	553 1 ASNB_ECOLI	P22106 escherichia
20	39	54.2	627 1 GLGB_BACSU	F39118 bacillus su
21	39	54.2	645 1 GLGB_BACAA	O81862 bacillus an
22	39	54.2	645 1 GLGB_BACAA	O81866 bacillus ce
23	39	54.2	799 1 SCA_DROME	P21520 drosophila
24	39	54.2	209 1 NUGM_DICDI	P22237 dictyostell
25	38	52.8	225 1 YC39_MENYA	O58566 methanococ
26	38	52.8	225 1 MYO2_ONCMY	O81206 oncorhynch
27	38	52.8	320 1 MMAL_DERPT	P08176 dermatophag
28	38	52.8	321 1 EUM1_EURWA	P25780 eurogilyphus
29	38	52.8	321 1 MMAL_DERFA	P16311 dermatophag
30	38	52.8	373 1 CYRG_CANFA	P40321 canis fami
31	38	52.8	481 1 MURC_TREDA	O83361 treponema p
32	38	52.8	572 1 FUB3_HUMAN	O96124 homo sapien
33	38	52.8	741 1 FIBA_CHICK	P14448 gallus gall

34	38	52.8	782 1 FIBA_RAT	P06399 rattus norv
35	38	52.8	866 1 FIBA_HUMAN	P02671 homo sapien
36	37.5	52.1	397 1 MP13_AMEAR	P27761 ambrosia ar
37	37.5	52.1	398 1 MP12_AMEAR	P27760 ambrosia ar
38	37.5	52.1	451 1 IRF4_HUMAN	O15306 homo sapien
39	37	51.4	318 1 YU1A_ECOLI	P24203 escherichia
40	37	51.4	338 1 MO2M_CAEEL	O18211 caenorhabd
41	37	51.4	345 1 YU44_ARCFU	O28335 archaeglob
42	37	51.4	374 1 ZNPD_WILMR	Q12723 willipsis
43	37	51.4	439 1 FGL2_HUMAN	Q14314 homo sapien
44	37	51.4	462 1 NFSL_DROME	O9y264 drosophila
45	37	51.4	503 1 ACP4_HUMAN	O9y264 homo sapien

## ALIGNMENTS

```

RESULT 1
ID YMS9_YEAST STANDARD PRT 457 AA.
AC Q03648;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical 52.2 kDa protein in RAR1-SC1 intergenic region.
GN YMR209C OR YMR261.03C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycos.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=5288C / AB972;
RX MEDLINE=97313268; PubMed=9169872;
RA Bowman S., Churchev C.M., Badcock K., Brown D., Chillingworth T.,
RA Connor R., Dedman K., Devlin K., Gentles S., Hamlin N., Hunt S.,
RA Jagels K., Lye G., Moule S., Odell C., Pearson D., Rajadream M.A.,
RA Rice P., Skellon J., Walsh S., Whitehead S., Barrall B.G.;
RA "The nucleotide sequence of Saccharomyces cerevisiae chromosome
RT XII."
RL Nature 387:90-93(1997).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements
CC or send an email to license@ebi.ac.uk).
CC -----
DR EMBL: Z49809; CAAB924.1; -.
DR PIR: S55091; S55091.
DR GenBank: J142884; -.
DR SCD: S0004822; YMR209C.
KW Hypothetical protein; Transmembrane.
FT TRANSMEM 1 21
FT TRANSMEM 250 270
FT TRANSMEM 457 522; MW: 52246 MW; B68737D1E58176E3 CRC64;
SQ SEQUENCE 457 AA; 52246 MW; B68737D1E58176E3 CRC64;

Query Match 58.3%; Score 42; DB 1; Length 457;
Best Local Similarity 63.6%; Pred. NO. 10;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

DB 2 RSNYDRSGWGY 12
330 RSNLDPLWGF 340

RESULT 2
ID YPV8_METTF STANDARD PRT 361 AA.
AC P29584;
DT 01-APR-1993 (Rel. 25, Created)

```

DT 01-APR-1993 (Rel. 25, Last annotation update)  
DT 01-APR-1993 (Rel. 25, Last annotation update)  
DE Hypothetical 40.7 kDa protein (ORF8).  
OS Methanobacterium thermoformicum.  
OG Plasmid pFV1.  
OC Archaeae; Euryarchaeota; Methanobacteria; Methanobacteriales;  
OC Methanobacteriaceae; Methanothermobacter.  
OX NCBI\_TaxID=145262;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=DSM 3848 / THF;  
RX MEDLINE=93126090; PubMed=1336177;  
RA Nolling J., van Eeden F.J.M., Eggen R.I.T., de Vos W.M.;  
RT "Modular organization of related Archaeal plasmids encoding different restriction-modification systems in Methanobacterium thermoformicum";  
RL Nucleic Acids Res. 20:6501-6507(1992).  
CC -1- SIMILARITY: TO THE CORRESPONDING ORF IN PFZ1.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: X68366; CAA48425.1; -  
CC PIR: S30313; S30313.  
KW Hypothetical protein; Plasmid.  
SQ SEQUENCE 361 AA; 40667 MW; 99CCADFE09E14680 CRC64;  
QY Query Match 56.9%; Score 41; DB 1; Length 361;  
Best Local Similarity 58.3%; Pred. No. 12;  
Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;  
DB 1 RRSYDRSGWDPY 12  
281 RRDLSRSMGPY 292  
RESULT 3  
GPMI\_BACME STANDARD; PRT; 511 AA.  
ID GPMI\_BACME  
AC P35167; O95655;  
DT 01-FEB-1994 (Rel. 28, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DE 28-FEB-2003 (Rel. 41, Last annotation update)  
DE 2,3-bisphosphoglycerate-independent phosphoglycerate mutase (EC 5.4.2.1) (Phosphoglyceromutase) (BPG-independent PGM) (iPGM).  
GN GPMI OR PGM.  
OS Bacillus megaterium.  
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
OX NCBI\_TaxID=1404;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 12872 / QMRDC B1551;  
RA Chander M., Setlow P., Laman E., Jedrejnas M.J.;  
RT "Structural studies on a 2,3-diphosphoglycerate independent phosphoglycerate mutase from Bacillus stearothermophilus";  
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE OF 1-232 FROM N.A.  
RC STRAIN=DSM 319;  
RX MEDLINE=93083995; PubMed=1452037;  
RA Schlaepfer B.S., Zuber H.;  
RT "Cloning and sequencing of the genes encoding glyceraldehyde-3-phosphate dehydrogenase, phosphoglycerate kinase and triosephosphate isomerase (gap operon) from mesophilic Bacillus megaterium: comparison with corresponding sequences from thermophilic Bacillus stearothermophilus";  
RL Gene 122:53-62(1992).  
CC -1- FUNCTION: Catalyzes the interconversion of 2-phosphoglycerate and

CC 3-phosphoglycerate (By similarity).  
CC -1- CATALYTIC ACTIVITY: 2-phospho-D-glycerate = 3-phospho-D-glycerate.  
CC -1- COFACTOR: Binds 2 manganese ions per subunit (By similarity).  
CC -1- PATHWAY: Glycolysis.  
CC -1- SUBUNIT: Monomer (By similarity).  
CC -1- SIMILARITY: Belongs to the BPG-independent phosphoglycerate mutase family.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: AF120090; AAP26327.1; -  
CC EMBL: M87647; AAA73205.1; -  
CC EMBL: M87648; AAA73208.1; -  
CC PIR: P00538; P00538.  
CC HAMAP: MF\_01038; -; 1.  
CC InterPro: IPR006124; Metalloenzyme.  
CC InterPro: IPR005995; Pgm\_bpd\_ind.  
CC Pfam: PF01676; Metalloenzyme\_1.  
CC PRODOM: PD004429; Pgm\_bpd\_ind; 1.  
CC TIGRFAWS: TIGR01307; pgm\_bpd\_ind; 1.  
CC Isoenzyme: Glycolysis; Metal-binding; Manganese.  
CC ACT\_SITE 62  
CC METAL 12  
CC METAL 12  
CC METAL 62  
CC METAL 403  
CC METAL 407  
CC METAL 407  
CC METAL 444  
CC METAL 444  
CC METAL 445  
CC METAL 462  
CC METAL 462  
CC CONFLICT 225  
CC FT 225  
CC SQ SEQUENCE 511 AA; 56563 MW; 7D3FAB44993B6C9 CRC64;  
QY Query Match 56.9%; Score 41; DB 1; Length 511;  
Best Local Similarity 41.7%; Pred. No. 17;  
Matches 5; Conservative 4; Mismatches 3; Indels 0; Gaps 0;  
DB 1 RRSYDRSGWDPY 12  
29 KKNFDFRWNEY 40  
RESULT 4  
YH02\_YEAST STANDARD; PRT; 602 AA.  
ID YH02\_YEAST  
AC P38867;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 01-FEB-1995 (Rel. 31, Last sequence update)  
DE 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Hypothetical 69.0 kDa protein in PPX1-RPS4B intergenic region. YHR202W.  
GN Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=S288C / AB972;  
RX MEDLINE=94378003; PubMed=8091229;  
RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J., Du Z., Favello A., Fulton L., Gattung S., Geisel C., Kirsten J., Kucaba T., Hillier L.W., Jier M., Johnston L., Langston Y., Latreille P., Louis E.J., Macri C., Mardis E., Meneses S., Mouser L., Nhan M., Rifkin L., Riles L., St Peter H., Trevaekis E., Vaughan K., Vignati D., Wilcox L., Woldman P., Waterston R., Wilson R., Vandin M.;  
RT "Complete nucleotide sequence of Saccharomyces cerevisiae chromosome

```

RT VILL."
CC Science 265:2077-2082(1994).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U00030; AAB68354.1; -.
DR PIR; S4676; S4676.
DR GERMOnline; 139520; -.
DR SGD; S0001245; YHR202W.
DR InterPro; IPR004843; M-peptidase.
DR Pfam; PF00149; Metallophos; 1.
KW Hypothetical protein.
SQ SEQUENCE 602 AA; 68998 MW; B9A69DAA3DD0A702 CRC64;

Query Match
Best Local Similarity 56.9%; Score 41; DB 1; Length 602;
Matches 5; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 2 RSNYDRSMGDI 12
DB 62 QNDYDADMGDF 72
:::|::|::|
:::|::|::|

RESULT 5
SYA DEIRA
ID SYA DEIRA STANDARD; PRT; 890 AA.
AC Q9RS27;
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Many-1-ORNA synthetase (SC 6.1.1.7) (Alanine--tRNA ligase) (AARS).
GN AARS OR DR2300.
OS Deinococcus radiodurans.
OC Bacteria; Deinococcus-Thermus; Deinococci; Deinococcales;
OC Deinococcaceae; Deinococcus.
OX NCBI_TaxID=1299;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RI / ATCC 13939 / DSM 20539 / NCIB 9279;
RX MEDLINE=2036896; PubMed=10567266;
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Hilt D.H., Gwin M.L., Nelson W.C., Richardson D.L.,
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Vamathevan J.J., Lam P., McDonald L., Uterback T., Zaleski C.,
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RT "Genome sequence of the radioresistant bacterium Deinococcus
RT radiodurans RI."
RT Science 286:1571-1577(1999).
CC -1- CATALYTIC ACTIVITY: ATP + L-alanine + tRNA(Ala) = AMP +
CC dihydrophosphate + L-alanyl-tRNA(Ala).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: Belongs to class-II aminoacyl-tRNA synthetase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE002062; AAF1848.1; -.
DR PIR; F75289; F75289.
DR TIGR; DR2300; -.
DR HAMAP; MF_00036; -. 1.

```

```

DR InterPro; IPR003156; DHHA1.
DR InterPro; IPR002318; tRNA-synt_2c.
DR InterPro; IPR006193; tRNA_synt_Ala.
DR Pfam; PF02272; DHHA1; 1.
DR Pfam; PF01411; tRNA-synt_2c; 1.
DR PRINTS; PRO0980; TRNASYNTHALA.
DR TIGRFAMs; TIGR00344; AARS; 1.
DR PROSITE; PS50860; AA tRNA LIGASE II ALA; 1.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW Complete proteome.
SQ SEQUENCE 890 AA; 95886 MW; FABE000AE34DAE2B CRC64;

Query Match
Best Local Similarity 56.2%; Score 40.5; DB 1; Length 890;
Matches 8; Conservative 1; Mismatches 3; Indels 1; Gaps 1;

QY 1 RSNY-DRSMGDI 12
DB 188 RGENYGDITWGDY 200
:::|::|::|
:::|::|::|

RESULT 6
VP28 METTF
ID VP28 METTF STANDARD; PRT; 361 AA.
AC P29585;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-APR-1993 (Rel. 25, Last annotation update)
DE Hypothetical 40.6 kDa protein (ORF8').
OS Methanobacterium thermoformicicum.
OC Plasmid pPZ1.
OC Archaea; Euryarchaeota; Methanobacteriales;
OC Methanobacteriaceae; Methanothermobacter.
OX NCBI_TaxID=145262;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DSM 3720 / Z-245;
RX MEDLINE=93126090; PubMed=1336177;
RA Noelling U., van Breden F.U.M., Eggen R.I.L., de Vos W.M.;
RT "Molecular organization of related Archaeal plasmids encoding different
RT restriction-modification systems in Methanobacterium
RT thermoformicicum."
RT Nucleic Acids Res. 20:6501-6507(1992).
CC -1- SIMILARITY: TO THE CORRESPONDING ORF IN pPZ1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X68367; GAA48439.1; -.
DR PIR; S30325; S30325.
KW Hypothetical protein; Plasmid.
SQ SEQUENCE 361 AA; 40565 MW; 5787C9D09CA6384F CRC64;

Query Match
Best Local Similarity 55.6%; Score 40; DB 1; Length 361;
Matches 7; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 RSNYDRSMGDI 12
DB 281 RRDGPRSMGDI 292
:::|::|::|
:::|::|::|

RESULT 7
RS1 MYCLE
ID RS1 MYCLE STANDARD; PRT; 481 AA.
AC P46836;
DT 01-NOV-1995 (Rel. 32, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)

```

DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE 30S ribosomal protein S1.  
GN RPSA OR M1332.  
OS Mycobacterium leprae.  
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
RX NCBI\_TaxID=1769;  
[1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=96059637; PubMed=7476188;  
RA Psihl H., Cole S.T.;  
RT "The Mycobacterium leprae genome: systematic sequence analysis  
RT identifies key catalytic enzymes, ATP-dependent transport systems and  
RT a novel *polA* locus associated with genomic variability";  
RL Mol. Microbiol. 16:909-919(1995).  
[2]  
RP SEQUENCE FROM N.A.  
RX STRAIN=TN;  
RX MEDLINE=21128732; PubMed=11234002;  
RA Cole S.T., Eigmeier K., Parkhill J., James K.D., Thomson N.R.,  
RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,  
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,  
RA Davies R.W., Devlin K., Duthoy S., Felwell T., Fraser A., Hamlin N.,  
RA Holroyd S., Hornsby T., Jagsels K., Lacroix C., Maclean J., Moule S.,  
RA Murphy L., Oliver K., Quail M.A., Rajadream M.A., Rutherford K.M.,  
RA Rutter S., Seeger K., Simon S., Simmonds W., Skelton J., Squares R.,  
RA Squires S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,  
RA Barrell B.G.;  
RT "Massive gene decay in the leprosy bacillus";  
RT Nature 409:1007-1011(2001).  
RL -1- FUNCTION: BINDS MRNA; THUS FACILITATING RECOGNITION OF THE  
INITIATION POINT. IT IS NEEDED TO TRANSLATE MRNA WITH A SHORT  
SHINE-DALGARNO (SD) PURINE-RICH SEQUENCE (BY SIMILARITY).  
CC -1- SIMILARITY: Belongs to the S1P family of ribosomal proteins.  
CC -1- SIMILARITY: Contains 4 S1 motif domains.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; Z46257; CA86365.1; -;  
DR EMBL; AL583921; CAC31763.1; -;  
DR PIR; H87081; H87081.  
DR PIR; S77660; S77660.  
DR HSSP; P05055; ISRO.  
DR Leptoma; ML1382; -;  
DR InterPro; IPR008994; Nucleic acid OB.  
DR InterPro; IPR000110; Ribosomal\_S1.  
DR InterPro; IPR003029; S1.  
DR Pfam; PF00575; S1; 4.  
DR PRINTS; PRO0681; RIBOSOMAL\_S1.  
DR SMART; SM00316; S1; 4.  
DR PROSITE; PS50126; S1; 4.  
KW Ribosomal protein; Repeat; RNA-binding; Complete proteome.  
FT DOMAIN 36 105 S1 MOTIF 1.  
FT DOMAIN 123 188 S1 MOTIF 2.  
FT DOMAIN 209 277 S1 MOTIF 3.  
FT DOMAIN 294 363 S1 MOTIF 4.  
FT CONFLICT 224 393 AFVUL -> CVCRS (IN REF. 1).  
FT CONFLICT 364 393 QANEDYEEDPAKYGADSDYDQNGYIP -> ADQNGIHL  
FT RGVPRGKVRVGGPATTSTRTSSSL (IN REF. 1).  
SQ SEQUENCE 481 AA; 53279 MW; 4F955F68DEA8E0B CRC64;  
Query Match 55.6%; Score 40; DB 1; Length 481;  
Best Local Similarity 50.0%; Pred. No. 24;  
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;  
Cy 1 RRSNYDRSMG 10  
:|:|:|:|

DB 105 KRAYERAWG 114  
RESULT 8  
RS1\_MYCTU STANDARD; PRT; 481 AA.  
AC 006147;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE 30S ribosomal protein S1.  
GN RPSA OR RV1630 OR M11666 OR MYCTO1B2.22.  
OS Mycobacterium tuberculosis.  
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
RX NCBI\_TaxID=1773;  
[1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=H37RV;  
RX MEDLINE=98295987; PubMed=9634230;  
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,  
RA Badoek K., Basham D., Brown D., Chillingworth T., Connor R.,  
RA Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holroyd S.,  
RA Hornsby T., Jagsels K., Krogh A., Maclean J., Moule S., Murphy L.,  
RA Oliver S., Osborne J., Quail M.A., Rajadream M.A., Rogers J.,  
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,  
RA Sultson J.E., Taylor K., Whitehead S., Barrell B.G.;  
RT "Deciphering the biology of Mycobacterium tuberculosis from the  
RT complete genome sequence";  
RT Nature 393:537-544(1998).  
[2]  
RP SEQUENCE FROM N.A.  
RX STRAIN=CDC 1551 / Oshkosh;  
RX MEDLINE=22206494; PubMed=12218036;  
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,  
RA Kolony J.F., Nelson W.C., Umayam L.A., Esmolaeva M., Salzberg S.L.,  
RA Delcher A., Utterback T., Weidman J., Kouri H., Gill J., Mikula A.,  
RA Bishtal W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;  
RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and  
RT laboratory strains";  
RT J. Bacteriol. 184:5479-5490(2002).  
RL -1- FUNCTION: BINDS MRNA; THUS FACILITATING RECOGNITION OF THE  
INITIATION POINT. IT IS NEEDED TO TRANSLATE MRNA WITH A SHORT  
SHINE-DALGARNO (SD) PURINE-RICH SEQUENCE (BY SIMILARITY).  
CC -1- SIMILARITY: Belongs to the S1P family of ribosomal proteins.  
CC -1- SIMILARITY: Contains 4 S1 motif domains.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; Z95554; CAB8883.1; -;  
DR EMBL; AE007030; AAK45936.1; -;  
DR PIR; D70559; D70559.  
DR HSSP; P05055; ISRO.  
DR TIGR; M11666; -;  
DR Tuberculist; RV1630; -;  
DR InterPro; IPR008994; Nucleic acid OB.  
DR InterPro; IPR000110; Ribosomal\_S1.  
DR InterPro; IPR003029; S1.  
DR Pfam; PF00575; S1; 4.  
DR PRINTS; PRO0681; RIBOSOMAL\_S1.  
DR SMART; SM00316; S1; 4.  
DR PROSITE; PS50126; S1; 4.  
KW Ribosomal protein; Repeat; RNA-binding; Complete proteome.  
FT DOMAIN 36 105 S1 MOTIF 1.  
FT DOMAIN 123 188 S1 MOTIF 2.

PT DOMAIN 209 277 S1 MOTIF 3.  
 FT DOMAIN 294 363 S1 MOTIF 4.  
 SQ SEQUENCE 481 AA; 53201 MW; FD73D8A5D051DBE1 CRC64;  
 Query Match 55.6%; Score 40; DB 1; Length 481;  
 Best Local Similarity 50.0%; Pred. No. 24;  
 Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 RNSYDRSMG 10  
 Db 105 KRAOYERANG 114

RESULT 9  
 TH1 MYCBO STANDARD; PRT; 277 AA.  
 ID TH1 MYCBO  
 AC PS9959;  
 DT 15-MAR-2004 (Rel. 43, Created)  
 DT 15-MAR-2004 (Rel. 43, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Putative thiosulfate sulfurtransferase 1 (EC 2.8.1.1) (Rhodanese-like protein 1).  
 GN CYSAL OR M80836C.  
 OS Mycobacterium bovis.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_TaxID=1765;  
 RX MEDLINE=22709107; PubMed=12788972;  
 RP STRAIN=AF2122/97;  
 RA Garnier T., Eigmeier K., Camus J.-C., Medina N., Mansoor H.,  
 RA Pryor W., Duthey S., Grondin S., Lacroix C., Mousseigne C., Simon S.,  
 RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,  
 RA Parkhill J., Barrett B.G., Cole S.T., Gordon S.V., Hewinson R.G.,  
 RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).  
 CC -1- FUNCTION: May be a sulfurtransferase involved in the formation of  
 CC thiosulfate (By similarity).  
 CC -1- CATALYTIC ACTIVITY: Thiosulfate + cyanide = sulfite + thiocyanate.  
 CC -1- SIMILARITY: Contains 2 rhodanese domains.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: BX248336; CAD93700.1; -;  
 DR PROSITE; PS00380; RHODANESE\_1; FALSE\_NEG.  
 DR PROSITE; PS00683; RHODANESE\_2; 1.  
 DR PROSITE; PS50206; RHODANESE\_3; 2.  
 KM Transferase; Repeat; Complete proteome.  
 FT DOMAIN 18 125 RHODANESE 1.  
 FT DOMAIN 154 274 RHODANESE 2.  
 FT ACT SITE 233 233 BY SIMILARITY.  
 FT SEQUENCE 277 AA; 31014 MW; AC37B715D9965A9 CRC64;  
 Query Match 54.2%; Score 39; DB 1; Length 277;  
 Best Local Similarity 66.7%; Pred. No. 20;  
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 4 NYDSRWGMDY 12  
 Db 257 NYDSRWGMDY 265

RESULT 10  
 TH1 MYCBO STANDARD; PRT; 277 AA.  
 ID TH1 MYCBO  
 AC Q50036;  
 DT 15-JUL-1998 (Rel. 36, Created)

DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Putative thiosulfate sulfurtransferase (EC 2.8.1.1) (Rhodanese-like protein).  
 GN CYSAL OR M2198.  
 OS Mycobacterium leprae.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_TaxID=1769;  
 RP STRAIN=TN;  
 RP SEQUENCE FROM N.A.  
 RP STRAIN=TN;  
 RX MEDLINE=21128732; PubMed=11234002;  
 RA Wheeler P.R., Eigmeier K., Parkhill J., James K.D., Thomson N.R.,  
 RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies R.M., Devlin K., Duthey S., Feltwell T., Fraser A., Hamlin N.,  
 RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,  
 RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,  
 RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,  
 RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,  
 RA Barrett B.G.,  
 RT "Massive gene decay in the leprosy bacillus."  
 RL Nature 409:1007-1011(2001).  
 CC -1- FUNCTION: May be a sulfurtransferase involved in the formation of  
 CC thiosulfate.  
 CC -1- CATALYTIC ACTIVITY: Thiosulfate + cyanide = sulfite + thiocyanate.  
 CC -1- SIMILARITY: Contains 2 rhodanese domains.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: U15182; AA623982.1; -;  
 DR EMBL; AL583924; CAC31153.1; -;  
 DR PIR; A87184; A87184.  
 DR Leprosy; M2198; -;  
 DR InterPro; IPR001763; Rhodanese-like.  
 DR InterPro; IPR001307; Rhodanese.  
 DR Pfam; PF00581; Rhodanese; 2.  
 DR SMART; SM00450; RHOD; 2.  
 DR PROSITE; PS00380; RHODANESE\_1; FALSE\_NEG.  
 DR PROSITE; PS00683; RHODANESE\_2; 1.  
 DR PROSITE; PS50206; RHODANESE\_3; 2.  
 KM Transferase; Complete proteome; Repeat.  
 FT DOMAIN 18 125 RHODANESE 1.  
 FT DOMAIN 154 274 RHODANESE 2.  
 FT ACT SITE 233 233 BY SIMILARITY.  
 FT SEQUENCE 277 AA; 31094 MW; 527E05C9FE9369D CRC64;  
 Query Match 54.2%; Score 39; DB 1; Length 277;  
 Best Local Similarity 66.7%; Pred. No. 20;  
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

RESULT 11  
 TH1 MYCBO STANDARD; PRT; 277 AA.  
 ID TH1 MYCBO  
 AC O05793;  
 DT 15-JUL-1998 (Rel. 36, Created)

DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DE 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Putative thiosulfate sulfurtransferase (EC 2.8.1.1) (Rhodanese-like  
protein).  
GN (CYSA1 OR CYSA OR RV3117 OR MT3199 OR MTGY164.27) AND (CYSA2 OR  
GN RV0815C OR MT0837 OR MTW043.07C).  
GN Mycobacterium tuberculosis.  
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.  
OX NCBI\_TaxID=1773;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=H37Rv;  
RX MEDLINE=96295987; PubMed=9634230;  
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,  
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
RA Davies K., Devlin K., Felwell T., Gentles S., Hamlin N., Holroyd S.,  
RA Horsbly T., Jagels K., Krogh A., McLean J., Moile S., Murphy L.,  
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,  
RA Sultun J.E., Taylor K., Whitehead S., Barrall B.G.;  
RT "Deciphering the biology of Mycobacterium tuberculosis from the  
RT complete genome sequence."  
RL Nature 393:537-544 (1998).  
[2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CDC 1551 / Oshkosh;  
RX MEDLINE=22206494; PubMed=12218036;  
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
RA Peterson J., Deboy R., Dodson R., Gwinn M., Hatt D., Hickey E.,  
RA Kolony J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,  
RA Delcher A., Utterback T., Weidman J., Kouri H., Gill J., Mukla A.,  
RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;  
RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and  
RT laboratory strains."  
RL J. Bacteriol. 184:5479-5490 (2002).  
CC -!- FUNCTION: May be a sulfotransferase involved in the formation of  
CC thiosulfate.  
CC -!- CATALYTIC ACTIVITY: Thiosulfate + cyanide = sulfite + thiocyanate.  
CC -!- SIMILARITY: Contains 2 rhodanese domains.  
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.ebi.ac.uk/announcement/>).  
CC or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).  
CC -----  
CC EMBL, Z95150; CAB08374.1; -  
DR EMBL, AL022004; CAAL7621.1; -  
DR EMBL, AE006973; AAK45079.1; -  
DR EMBL, AE007136; AAK47539.1; -  
DR PIR, G70809; G70809.  
DR HSP, P00586; IRHS.  
DR TIGR, MT0837; -  
DR TIGR, MT3199; -  
DR Tuberculat, RV0815C; -  
DR InterPro, IPR001763; Rhodanese-like.  
DR InterPro, IPR001307; Rhodanese.  
DR Pfam, PF00581; Rhodanese; 2.  
DR SMART, SM00450; RHOD; 2.  
DR PROSITE, PS00380; RHODANES\_1; FALSE\_NEG.  
DR PROSITE, PS00683; RHODANES\_2; 1.  
DR PROSITE, PS50206; RHODANES\_3; 2.  
DR Transferase; Repeat; Complete proteome.  
KW DOMAIN 18 125 RHODANES\_1.  
FT ACT\_SITE 154 274 RHODANES\_2.  
FT ACT\_SITE 233 233 BY SIMILARITY.  
SQ SEQUENCE 277 AA; 31014 MW; AC378715D959565A9 CRC64;  
54.2%; Score 39; DB 1; Length 277;

```

CY      4 NYDRSGMDY 12
      ||| ||| :|
      257 NYDGSWTXY 265

Db      257 NYDGSWTXY 265

RESULT 12
THTR_SACER
ID THTR_SACER STANDARD; PRT; 281 AA.
AC P16385;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Putative thiosulfate sulfurtransferase (EC 2.8.1.1) (Rhodanese-like
DE protein).
GN CysA.
OS Saccharopolyspora erythraea (Streptomyces erythraeus).
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Pseudonocardineae; Pseudonocardiaceae; Saccharopolyspora.
OX NCBI_Taxid=1836;
[1]
AN
AP
BP SEQUENCE FROM N.A., AND SEQUENCE OF 3-16.
RC STRAIN=WM22;
RX MEDLINE=90094240; Pubmed=2294090;
RA Donadio S., Shalite A., Hutchinson C.R.;
RT "Disruption of a rhodanese-like gene results in cysteine auxotrophy in
RT Saccharopolyspora erythraea."
RL J. Bacteriol. 172:350-360(1990).
CC -1 PUNCTON: May be a sulfotransferase involved in the formation of
CC thiosulfate.
CC -1 CATALYTIC ACTIVITY: Thiosulfate + cyanide = sulfite + thiocyanate.
CC -1 DOMAIN: THE STRUCTURE CONSISTS OF 2 DOMAINS OF VERY SIMILAR
CC CONFORMATION, SUGGESTING A COMMON EVOLUTIONARY ORIGIN. HOWEVER,
CC THE SEQUENCES OF THE 2 DOMAINS ARE VERY DIFFERENT.
CC -1 SIMILARITY: Contains 2 rhodanese domains.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC
CC EMBL; M29612; AA88935.1; .
CC InterPro; IPR001763; Rhodanese-like.
CC InterPro; IPR001307; Rhodanese.
CC Pfam; PF00581; Rhodanese; 2.
CC SMART; SMO0450; RHOD; 2.
CC DR PROSITE; PS00380; RHODANESE_1; 1.
CC DR PROSITE; PS00683; RHODANESE_2; 1.
CC DR PROSITE; PS50206; RHODANESE_3; 2.
CC KW Transferase; Repeat.
CC FT DOMAIN 18 125 RHODANSE 1.
CC FT DOMAIN 154 274 RHODANSE 2.
CC FT ACT SITE 233 233 BY SIMILARITY
CC SO SEQUENCE 281 AA; 31424 MW; 92CAB84E6926AE CRC64;

Query Match 54.2%; Score 39; DB 1; Length 281;
Best Local Similarity 66.7%; Pred. No. 20;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

CY      4 NYDRSGMDY 12
      ||| ||| :|
      257 NYDGSWTXY 265

Db      257 NYDGSWTXY 265

RESULT 13
FGL1_HUMAN
ID FGL1_HUMAN STANDARD; PRT; 312 AA.
AC Q08830; Q96KW6; Q96QW6;

```



DR 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Fibrinogen-like protein 1 precursor (hepatocyte-derived fibrinogen-  
 related protein 1) (HFRP-1) (Hepasocin) (HP-041).  
 GN FGL1 OR HFRSP1.  
 OS Homo sapiens (human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Liver;  
 RX MEDLINE=93290661; PubMed=8390249;  
 RA Yamamoto T., Gotch M., Sasaki H., Terada M., Kitajima M.,  
 RA Hirohashi S.;  
 RT "Molecular cloning and initial characterization of a novel  
 RT fibrinogen-related gene, HFRP-1.";  
 RL Biochem. Biophys. Res. Commun. 193:681-687(1993).  
 RN [2]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RC TISSUE=Liver;  
 RX MEDLINE=21363035; PubMed=11470158;  
 RA Hara H., Yoshimura H., Uchida S., Toyoda Y., Aoki M., Sakai Y.,  
 RA Morimoto S., Shiohara K.;  
 RT "Molecular cloning and functional expression analysis of a cDNA for  
 RT human hepasocin, a liver-specific protein with hepatocyte mitogenic  
 RT activity.";  
 RL Biochim. Biophys. Acta 1520:45-53(2001).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Liver;  
 RX MEDLINE=92368257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Krausner R.D., Collins F.S., Wagner L.D., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heish F.,  
 RA Diatchenko L., Maustria K., Farmer A.A., Rubin G.M., Hong J.,  
 RA Stupelstein M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Udell T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loughran N.A., Peters G.J., Abramson R.D., Mallary S.J.,  
 RA Bosak S.A., McKean P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W.,  
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Heintz B., Kettelman K., Madan A., Rodriguez S., Sanchez A.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,  
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length  
 RT human and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16699-16903(2002).  
 CC -1- FUNCTION: Has hepatocyte mitogenic activity.  
 CC -1- SUBUNIT: Homodimer (Probable).  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- TISSUE SPECIFICITY: Liver specific.  
 CC -1- SIMILARITY: Contains 1 fibrinogen C-terminal domain.  
 CC -1- This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: D14446; BA03336.1; -  
 DR EMBL: D87342; BAB70690.1; -  
 DR EMBL: BC007047; AA07047.1; -  
 DR PIR: JN0596; JN0596.  
 DR HSPSP: P02671; 1E2D.  
 DR Genew: HGNC:3695; FGL1.

DR MIM: 605776; -  
 DR GO: GO:0005577; C:fibrinogen complex; TAS.  
 DR InterPro: IPR002181; Fibrinogen\_C.  
 DR Pfam: PF00147; fibrinogen\_C.1.  
 DR SMART: SM00186; FBG.1.  
 DR PROSITE: PS00514; FIBRIN\_AG\_C\_DOMAIN; 1.  
 KW Signal.  
 FT SIGNAL 1 22  
 FT CHAIN 23 312  
 FT DOMAIN 78 305  
 FT DISULFID 26 26  
 FT DISULFID 83 112  
 FT DISULFID 248 261  
 FT CONFLICT 15 15  
 FT CONFLICT 69 69  
 FT CONFLICT 72 72  
 FT CONFLICT 105 105  
 SQ SEQUENCE 312 AA; 36391 MW; 26BC82124E6660C2 CRC64;  
 Query Match 54.2%; Score 39; DB 1; Length 312;  
 Best Local Similarity 50.0%; Pred. No. 23;  
 Matches 8; Conservative 2; Mismatches 2; Indels 4; Gaps 1;  
 QY 1 RRS----NYDRSGWGDY 12  
 Db 125 RRSQSGENFNKGMWDY 140  
 RESULT 14  
 CYRG MOUSE STANDARD; PRT: 369 AA.  
 ID CYRG MOUSE  
 AC P34802;  
 DT 01-FEB-1994 (Rel. 28, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Cytokine receptor common gamma chain precursor (Gamma-C)  
 DE (interleukin-2 receptor gamma chain) (IL-2R gamma chain) (P64).  
 GN IL2RG.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93277575; PubMed=8503926;  
 RA Kumaki S., Kondo M., Takeshita T., Asao H., Nakamura M., Sugamura K.;  
 RT "Cloning of the mouse interleukin 2 receptor gamma chain:  
 RT demonstration of functional differences between the mouse and human  
 RT receptors.";  
 RL Biochem. Biophys. Res. Commun. 193:356-363(1993).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CB6/CA;  
 RX MEDLINE=93391374; PubMed=8378320;  
 RA Cao X., Kozak C.A., Liu Y.J., Noguchi M., O'Connell E., Leonard W.J.;  
 RT "Characterization of cDNAs encoding the murine interleukin 2 receptor  
 RT (IL-2R) gamma chain: chromosomal mapping and tissue specificity of  
 RT IL-2R gamma chain expression.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 90:8464-8468(1993).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93166191; PubMed=8359699;  
 RA Kobayashi N., Nakagawa S., Minami Y., Taniguchi T., Kono T.;  
 RT "Cloning and sequencing of the cDNA encoding a mouse IL-2 receptor  
 RT gamma.";  
 RL Gene 130:303-304(1993).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=95104285; PubMed=7805729;  
 RA Disanto J.P., Certain S., Wilson A., Macdonald H.R., Ayner P.,  
 RA Fischer A., de Saint Basile G.;  
 RT "The murine interleukin-2 receptor gamma chain gene: organization,  
 RT chromosomal localization and expression in the adult thymus.";

RL Eur. J. Immunol. 24:3014-3018(1994).  
 [5]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-B6.S.  
 RX MEDLINE=9634745; PubMed=8750189;  
 RA Chu R.K., Droll A., Cooper D.L., Dougherty S.T., Dirks J.F.,  
 RA Dougherty G.J.  
 RT "Molecular mechanisms regulating the hyaluronan binding activity of  
 RT the adhesion protein CD44.";  
 RL J. Neurosci. 26:231-239(1995).  
 [6]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-FVB/N; TISSUE=Salivary gland;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Ditchenko L., Marzetta K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stepien M., Soares W.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Utschi T.B., Tomshy S., Carninci P., Prange C.,  
 RA Rata S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
 RA Bosak S.A., McEwen P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W.,  
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whitting M., Madan A., Young A.C., Shevchenko Y., Boulford G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickinson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S.N., Krzywinski M.I., Skalski U., Smalley D.E.,  
 RA Schenker A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length  
 RT human and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 CC -1- FUNCTION: Common subunit for the receptors for a variety of  
 CC interleukins.  
 CC -1- SUBUNIT: The gamma chain is common to the IL2, IL4, IL7, IL21 and  
 CC probably also the IL3 receptors.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- SIMILARITY: Belongs to the type I cytokine family of receptors.  
 CC -1- SIMILARITY: Contains 1 fibronectin type III domain.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 -----  
 DR EMBL; D13821; BAA02974.1; -;  
 DR EMBL; U21795; BAA04779.1; -;  
 DR EMBL; D13665; BAA02760.1; -;  
 DR EMBL; L20048; AAA32886.1; -;  
 DR EMBL; S75852; AAA32904.1; -;  
 DR EMBL; S75844; AAA32904.1; JOINED.  
 DR EMBL; S75845; AAA32904.1; JOINED.  
 DR EMBL; S75847; AAA32904.1; JOINED.  
 DR EMBL; S75848; AAA32904.1; JOINED.  
 DR EMBL; S75849; AAA32904.1; JOINED.  
 DR EMBL; S75850; AAA32904.1; JOINED.  
 DR EMBL; S75851; AAA32904.1; JOINED.  
 DR EMBL; X75337; CAAS3085.1; -;  
 DR EMBL; BC014720; AAH14720.1; -;  
 DR PIR; I49280; I49280.  
 DR HSSP; P31785; I11M.  
 DR MGI; 96551; I1279.  
 DR InterPro; IPR002996; CRIA.  
 DR InterPro; IPR008957; FN\_III-like.  
 DR InterPro; IPR003961; FN\_III.  
 DR InterPro; IPR003531; Hemitopoptn\_S\_Fl.  
 DR Pfam; PF00041; fn3; 1.

DR SMART; SMO0060; FN3; 1.  
 DR PROSITE; PS01355; HEMATOPO. REC. S. Fl; 1.  
 KM Receptor; Transmembrane; Glycoprotein; Signal.  
 FT SIGNAL 1 22  
 FT CHAIN 23 369  
 FT DOMAIN 23 263  
 FT TRANSMEM 264 284  
 FT DOMAIN 285 369  
 FT DOMAIN 151 250  
 FT DISULFID 62 72  
 FT DISULFID 102 115  
 FT CARBOHYD 71 71  
 FT CARBOHYD 75 75  
 FT CARBOHYD 84 84  
 FT CARBOHYD 96 96  
 FT CARBOHYD 159 159  
 FT CARBOHYD 164 164  
 SQ SEQUENCE 369 AA; 42241 MW; C82D5AB459077AC7 CMC64;  
 Query Match 54.2%; Score 39; DB 1; Length 369;  
 Best Local Similarity 87.5%; Pred. No. 27;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 2 RSNDPSW 9  
 Db 131 RSNDRSW 138  
 RESULT 15  
 ID IRF4 MOUSE  
 AC 064287; 060802; STANDARD; PRT; 450 AA.  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Interferon regulatory factor 4 (IRF-4) (Lymphocyte specific interferon  
 DE regulatory factor) (LIRF) (NF-E2M3) (PU.1 interaction partner)  
 DE (transcriptional activator PIP).  
 GN IRF4 OR SDFP.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI Taxid=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BALB/c;  
 RX MEDLINE=95317607; PubMed=7797077;  
 RA Eisenbeis C.F., Singh H., Storb U.;  
 RT "PIP, a novel IRF family member, is a lymphoid-specific,  
 RT PU.1-dependent transcriptional activator.";  
 RL Genes Dev. 9:1377-1387(1995).  
 [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-C57BL/6; and 129/SvJ; TISSUE=Spleen;  
 RX MEDLINE=9534564; PubMed=7541907;  
 RA Matsuyama T., Grossman A., Mitrucker H.-N., Siderovski D.P.,  
 RA Kiefer F., Kawakami T., Richardson C.D., Taniguchi T., Yoshinaga S.K.,  
 RA Mak T.W.;  
 RT "Molecular cloning of LSIRF, a lymphoid-specific member of the  
 RT interferon regulatory factor family that binds the interferon-  
 RT stimulated response element (ISRE)."  
 RL Nucleic Acids Res. 23:2127-2136(1995).  
 CC -1- FUNCTION: TRANSCRIPTIONAL ACTIVATOR. BINDS TO THE INTERFERON-  
 CC STIMULATED RESPONSE ELEMENT (ISRE) OF THE MHC CLASS I PROMOTER.  
 CC BINDS THE IMMUGLOBULIN LAMBDA LIGHT CHAIN ENHANCER, TOGETHER  
 CC WITH PU.1. PROBABLY PLAYS A ROLE IN ISRE-TARGETED SIGNAL  
 CC TRANSDUCTION MECHANISMS SPECIFIC TO LYMPHOID CELLS.  
 CC  
 CC -1- SUBCELLULAR LOCATION: Nucleus.  
 CC  
 CC -1- ALTERNATIVE PRODUCTS:  
 CC Event-Alternative splicing; Named isoforms=2;  
 CC Name=1;  
 CC IsoId=064287-1; Sequence=Displayed;  
 CC Name=2;

```

CC      IscId=Q64287-2; Sequence=VSP_002756;
CC      -1- TISSUE SPECIFICITY: LYMPHOID CELLS.
CC      -1- INDUCTION: Not induced by interferons.
CC      -1- SIMILARITY: Belongs to the IRF family.
CC
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL; U34307; AAA75283.1; -
CC      EMBL; U11692; AAA75309.1; -
CC      EMBL; U20949; AAA75316.1; -
CC      EMBL; U20949; AAA75317.1; -
CC      PIR; S57837; S57837.
CC      HSSP; P23906; 2IRF.
CC      TRANSFAC; T01572; -
CC      MGD; MGI:1096873; Irfa.
CC      InterPro; IPR001346; IRF.
CC      InterPro; IPR008984; SMAD_FHA.
CC      Pfam; PF00605; IRF_1.
CC      PRINTS; PR00267; INTERNRREGCT.
CC      PRODOM; PD002355; IRF_1.
CC      SMART; SM00348; IRF_1.
CC      PROSITE; PS00601; IRF_1.
CC      KW      Transcription regulation; DNA-binding; Nuclear protein; Activator;
CC      Alternative splicing.
CC      FT      DNA BIND          TRYPTOPHAN PENTAD REPEAT.
CC      FT      23              125      Missing (in isoform 2).
CC      FT      VARSPLIC 165      165      /FTId=VSP_002756.
CC      SQ      SEQUENCE 450 AA; 51577 MW; 5FD94CA6C453869C CRC64;
OY      5 YDRSMGDPY 12
DB      174 HDRSMRDPY 181

```

Search completed: October 5, 2004, 08:01:48  
 Job time : 3.07407 secs

```

Query Match          54.2%; Score 39; DB 1; Length 450;
Best Local Similarity 75.0%; Pred. No. 33;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```

This Page Blank (uspio)

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2004, 07:34:11 : Search time 6.35185 Seconds  
(without alignments)  
596.081 Million cell updates/sec

Title: US-09-805-290A-14  
Perfect score: 72  
Sequence: 1 RSNVDRSGMDY 12

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315318202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

SPREMBL\_25: \*  
1: sp\_archaea: \*  
2: sp\_bacteria: \*  
3: sp\_fungi: \*  
4: sp\_human: \*  
5: sp\_invertebrate: \*  
6: sp\_mammal: \*  
7: sp\_mhc: \*  
8: sp\_organelle: \*  
9: sp\_phage: \*  
10: sp\_plant: \*  
11: sp\_rodent: \*  
12: sp\_virus: \*  
13: sp\_vertebrate: \*  
14: sp\_unclassified: \*  
15: sp\_virus: \*  
16: sp\_bacteriophage: \*  
17: sp\_archaeop: \*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	46	63.9	332	10 Q8SB49	Q8SB49 oryza sativ
2	46	63.9	332	10 Q7XF61	Q7XF61 oryza sativ
3	46	63.9	491	16 Q8G5L7	Q8G5L7 bifidobacte
4	46	63.9	730	16 Q8P9U7	Q8P9U7 xanthomonas
5	46	63.9	906	16 Q8P1M6	Q8P1M6 xanthomonas
6	43	59.7	91	16 Q87TN7	Q87TN7 clostridium
7	43	59.7	104	16 Q87TU3	Q87TU3 leptospira
8	43	59.7	224	11 Q83409	Q83409 rattus norv
9	43	59.7	1057	16 Q8A3Y6	Q8A3Y6 bacteroides
10	42.5	59.0	370	5 Q812W2	Q812W2 plasmodium
11	42	58.3	37	16 Q87UR0	Q87UR0 pseudomonas
12	42	58.3	251	5 Q81Q23	Q81Q23 drosophila
13	42	58.3	356	10 Q7XQV1	Q7XQV1 oryza sativ
14	42	58.3	358	5 Q8VX04	Q8VX04 drosophila
15	42	58.3	355	10 Q81NE3	Q81NE3 oryza sativ
16	42	58.3	427	10 Q94H68	Q94H68 oryza sativ

17	42	58.3	557	10 Q7X6W4	Q7X6W4 oryza sativ
18	42	58.3	665	10 Q7XMR7	Q7XMR7 oryza sativ
19	42	58.3	666	2 Q54272	Q54272 staphylococ
20	42	58.3	666	2 P96018	P96018 staphylococ
21	42	58.3	666	2 Q54277	Q54277 staphylococ
22	42	58.3	678	10 Q94168	Q94168 oryza sativ
23	42	58.3	722	10 Q7XPG0	Q7XPG0 oryza sativ
24	42	58.3	785	10 Q8G5S1	Q8G5S1 oryza sativ
25	42	58.3	806	10 Q8G2G7	Q8G2G7 oryza sativ
26	42	58.3	823	10 Q8S848	Q8S848 oryza sativ
27	42	58.3	823	10 Q84S82	Q84S82 oryza sativ
28	42	58.3	829	10 Q7XMS6	Q7XMS6 oryza sativ
29	42	58.3	839	10 Q7XMS6	Q7XMS6 oryza sativ
30	42	58.3	905	10 Q7X8G3	Q7X8G3 oryza sativ
31	42	58.3	3657	4 Q96Q15	Q96Q15 homo sapien
32	42	58.3	98	16 Q8Y0E2	Q8Y0E2 raietonia s
33	41	56.9	107	16 Q8Y0E2	Q8Y0E2 raietonia s
34	41	56.9	152	16 Q9A2T9	Q9A2T9 caulobacter
35	41	56.9	227	11 Q8X583	Q8X583 rattus norv
36	41	56.9	314	11 Q8YC25	Q8YC25 mus musculu
37	41	56.9	415	9 Q8SDX8	Q8SDX8 staphylococ
38	41	56.9	415	9 Q8NVF8	Q8NVF8 staphylococ
39	41	56.9	448	16 Q888E8	Q888E8 rhizobium 1
40	41	56.9	462	5 Q9U7P2	Q9U7P2 eufoleculi
41	41	56.9	550	16 Q98Q90	Q98Q90 mycoplasma
42	41	56.9	654	10 Q7XN46	Q7XN46 oryza sativ
43	41	56.9	714	10 Q8RZ31	Q8RZ31 oryza sativ
44	41	56.9	835	10 Q8LI08	Q8LI08 oryza sativ
45	41	56.9	859	10 Q7XWB4	Q7XWB4 oryza sativ

## ALIGNMENTS

RESULT 1					
Q8SB49		PRELIMINARY;	PRT;	332 AA.	
AC Q8SB49					
DT 01-JUN-2002 (TREMBlrel. 21, Created)					
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)					
DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)					
DE Putative polypeptide.					
GN OSJNB0091009.10.					
OS Oryza sativa (Rice).					
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;					
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;					
OC Erihartoideae; Oryzaceae; Oryza.					
OC NCBI_TaxID=4530;					
RN [1]					
RP SEQUENCE FROM N.A.					
RA Wang R.A., Yu Y., Soderlund C., Chen M., Kim H.-R., Rambo T.,					
RA Sasaki C., Henry D., Oates R., Simmons J.;					
RT "Rice Genomic Sequence.";					
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.					
DR EMBL; AC091732; AL177159.1; -.					
DR Gramene; Q8SB49; -.					
KW Polypeptide.					
SC SEQUENCE 332 AA; 36025 MW; 06B36509F415AB5C CRC64;					
QY	Query Match	63.9%	Score 46;	DB 10;	Length 332;
Db	Best Local Similarity	58.3%	Prod. No. 13;		
	Matches 7;	Conservative 2;	Mismatches 3;	Indels 0;	Gaps 0;
	1 RSNVDRSGMDY 12				
	98 RSKNFDKRWGDM 109				
RESULT 2					
Q7XF61		PRELIMINARY;	PRT;	332 AA.	
ID Q7XF61					
AC Q7XF61					
DT 01-OCT-2003 (TREMBlrel. 25, Created)					

[illegible]

OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;  
 OC Xanthomonadaceae; Xanthomonas.  
 OX NCBI\_TaxId=92829;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=306 / ATCC 13902 / XV 101;  
 RX MEDLINE=22022145; PubMed=12024217;  
 RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furian L.R.,  
 RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,  
 RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,  
 RA Camarotte G., Camarvan F., Cardozo J., Chamberg F., Clapina L.P.,  
 RA Ciccarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorry H.,  
 RA Faria U.B., Ferreira A.U.S., Ferreira R.C.C., Ferro M.I.T.,  
 RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,  
 RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,  
 RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,  
 RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,  
 RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,  
 RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,  
 RA Spindola L.A.F., Takita W.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,  
 RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,  
 RA Setubal J.C., Kitajima J.P.;  
 RA "Comparison of the genomes of two Xanthomonas pathogens with differing  
 RT host specificities.";  
 RL Nature 417:459-463(2002).  
 DR EMBL; AE011808; AM36633.1; -  
 DR GO; GO:0016020; C:membrane; IEA.  
 DR GO; GO:0004872; F:receptor activity; IEA.  
 DR GO; GO:0005215; F:transporter activity; IEA.  
 DR GO; GO:0006810; P:transport; IEA.  
 DR InterPro; IPRO00531; TonB\_boxC.  
 DR Pfam; PF00593; TonB\_deg\_Reg; 1.  
 DR PROSITE; PS00430; TONB\_DEPENDENT\_REC\_1; 1.  
 KW Receptor; Complete proteome.  
 DR PROSITE; PS00430; TONB\_DEPENDENT\_REC\_1; 1.  
 KW Receptor; Complete proteome.  
 SQ SEQUENCE 906 AA; 99417 MW; 191B59A91992185 CRC64;  
 QY 2 RSNYDRSGWDY 12  
 Db 307 RSDYDFSMHDY 317  
 Query Match 63.9%; Score 46; DB 16; Length 906;  
 Best Local Similarity 72.7%; Pred. No. 37;  
 Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 RESULT 6  
 Q97TN7 PRELIMINARY; PRT; 91 AA.  
 AC Q97TN7;  
 DT 01-OCT-2001 (TREMBlrel. 18, Created)  
 DT 01-OCT-2001 (TREMBlrel. 18, Last sequence update)  
 DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)  
 DE Hypothetical protein CAP0061.  
 GN CAP0061.  
 OS Clostridium acetobutylicum.  
 OG Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;  
 OC Clostridium.  
 OC Clostridium.  
 NCBI\_TaxId=1488;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;  
 RX MEDLINE=21359325; PubMed=11466286;  
 RA Neill J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,  
 RA Gibson R., Lee H.M., Dubois J., Qiu J., Hilti J., Wolf Y.I.,  
 RA Tatunov R.L., Sababe F., Doucette-Stamm L., Souaille P., Daly M.J.,  
 RA Bennett G.N., Koonin E.V., Smith D.R.,  
 RA "Genome sequence and comparative analysis of the solvent-producing  
 RT bacterium Clostridium acetobutylicum.";  
 RL J. Bacteriol. 183:4823-4838(2001).  
 DR EMBL; AE001438; AAK76807.1; -  
 DR GO; GO:0046821; C:extrachromosomal DNA; IEA.  
 DR GO; GO:0003700; F:transcription factor activity; IEA.

DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.  
 DR InterPro; IPRO00792; HTH\_LuxR.  
 KW Plasmid; Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 91 AA; 10974 MW; 09EB513233E67F4D CRC64;  
 Query Match 59.7%; Score 43; DB 16; Length 91;  
 Best Local Similarity 50.0%; Pred. No. 9.8;  
 Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 RSNYDRSGWDY 12  
 Db 31 RKSNSKXKAMGEY 42  
 RESULT 7  
 Q97U3 PRELIMINARY; PRT; 104 AA.  
 AC Q97U3;  
 DT 01-MAR-2003 (TREMBlrel. 23, Created)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)  
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
 DE Ribosomal protein L21.  
 DE RPLU OR LA0848.  
 OS Leptospira interrogans.  
 OC Bacteria; Spirochaetes; Spirochaetales; Leptospiraceae; Leptospira.  
 OX NCBI\_TaxId=173;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=56501 / Serogroup Icterohemorrhagiae / Serovar lai;  
 RA Ren S.;  
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AE011270; AA48048.1; -  
 DR GO; GO:0005622; C:intracellular; IEA.  
 DR GO; GO:0005840; C:ribosome; IEA.  
 DR GO; GO:0003723; F:RNA binding; IEA.  
 DR GO; GO:0003735; F:structural constituent of ribosome; IEA.  
 DR GO; GO:0006412; P:protein biosynthesis; IEA.  
 DR InterPro; IPRO01787; Ribosomal\_L21p.  
 DR Pfam; PF00829; Ribosomal\_L21p; 1.  
 DR ProDom; PD003604; Ribosomal\_L21p; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 104 AA; 11879 MW; 66B90159128B59D CRC64;  
 QY 1 RSNYDRSGWG 10  
 Db 79 RKNYOKXWG 88  
 Query Match 59.7%; Score 43; DB 16; Length 104;  
 Best Local Similarity 60.0%; Pred. No. 11;  
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
 RESULT 8  
 Q63409 PRELIMINARY; PRT; 224 AA.  
 AC Q63409;  
 DT 01-NOV-1996 (TREMBlrel. 01, Created)  
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
 DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)  
 DE Olt protein (Fragment).  
 GN OLT.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxId=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Sprague-Dawley;  
 RA Kerr S.M., Taggart M.H., Cooke H.J.;  
 RA "Unusually high sequence variation in rodent Olt genes.";  
 RT Submitted (Aug-1996) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; X99901; CAA68173.1; -  
 FT NON\_TER 1

FT NON TER 224 224  
 SEQ SEQUENCE 224 AA; 26983 MW; 962E67422790F537 CRC64;  
 Query Match 59.7%; Score 43; DB 11; Length 224;  
 Best Local Similarity 77.8%; Pred. No. 26;  
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 RSNYDRSGW 9  
 DB 200 RRDYDRSGW 208

## RESULT 9

Q8A3Y6 PRELIMINARY; PRT; 1057 AA.  
 AC Q8A3Y6; 01-JUN-2003 (TRENBLREL. 24, Created)  
 DT 01-JUN-2003 (TRENBLREL. 24, Last sequence update)  
 DT 01-OCT-2003 (TRENBLREL. 25, Last annotation update)  
 DE Putative outer membrane protein, probably involved in nutrient binding.  
 DE BT2818.  
 GN Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;  
 OC Bacteroidaceae; Bacteroides.  
 CX NCBI\_TaxID=818;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=VPI-5482 / ATCC 29148;  
 RX MEDLINE=22550858; PubMed=12663928;  
 RA Xu J., Bjursell M.K., Himrod J., Deng S., Carmichael L.K., Chiang H.C., Hooper L.V., Gordon J.I.;  
 RT "A genomic view of the human-Bacteroides thetaiotaomicron symbiosis.";  
 RL Science 289:2074-2076(2003).  
 DR EMBL; AE016937; AA077924.1;  
 DR GO; GO:0016020; C:membrane; IEA.  
 DR GO; GO:0004872; F:receptor activity; IEA.  
 DR GO; GO:0005215; F:transporter activity; IEA.  
 DR GO; GO:0006810; P:transport; IEA.  
 DR InterPro; IPR008969; Carboxypept\_reg.  
 DR InterPro; IPR00531; TonB\_boxC.  
 DR Pfam; PF00593; TonB\_dep\_Reg; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 1057 AA; 117916 MW; DA28E63C44DF05C CRC64;  
 Query Match 59.7%; Score 43; DB 16; Length 1057;  
 Best Local Similarity 63.6%; Pred. No. 1.4e+02;  
 Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 RSNYDRSGW 12  
 DB 562 RINDYRTFGDH 572

## RESULT 10

Q812W2 PRELIMINARY; PRT; 370 AA.  
 AC Q812W2; 01-MAR-2003 (TRENBLREL. 23, Created)  
 DT 01-MAR-2003 (TRENBLREL. 23, Last sequence update)  
 DT 01-OCT-2003 (TRENBLREL. 25, Last annotation update)  
 DE DnaJ-like molecular chaperone protein, putative.  
 DE PFT0935M.  
 GN Plasmodium falciparum (isolate 3D7).  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 CX NCBI\_TaxID=36329;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22255708; PubMed=12368867;  
 RA Hall N., Bain A., Berriman M., Churcher C., Harris B., Harris D., Mungall K., Bowman S., Atkin R., Baker S., Barton A., Brooks K., Buckee C.O., Burrows C., Cherevach I., Chillingworth C., Chillingworth T., Christodoulou Z., Clark L., Clark R., Corton C.,

RA Cronin A., Davies R., Davis P., Dear P., Dearden F., Doggett J., RA Felwell T., Goble A., Goodhead I., Gwilliam R., Hamlin N., Hance Z., RA Harper D., Hauser H., Hornsby T., Holroyd S., Horrocks P., RA Humphrey S., Jagels K., James K.D., Johnson D., Kerkhoun A., RA Knights A., Konfortov B., Kyes S., Larke N., Lawson D., Lennard N., RA Line A., Maddison M., McLean J., Mooney P., Moule S., Murphy L., RA Oliver K., Ormond D., Price C., Quail M.A., Rabinowitsch E., RA Rajadream M.A., Rutter S., Rutherford K.M., Sanders M., Simmonds M., RA Seeger K., Sharp S., Smith R., Squares R., Squares K., Taylor K., Tivey A., Twinn L., Whitehead S., Woodward J., RA Sulston J.S., Craig A., Newbold C., Barrel B.G.;  
 RT "Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13.";  
 RL Nature 419:527-531(2002).  
 DR EMBL; AL928357; CAD51872.1;  
 DR GO; GO:0003754; F:chaperone activity; IEA.  
 DR InterPro; IPR001623; DnaJ\_N.  
 DR InterPro; IPR003095; Hsp\_DnaJ.  
 DR Pfam; PF00226; DnaJ; 1.  
 DR PRINTS; PR00625; DnaJPROTEIN.  
 DR SMART; SM00271; DnaJ; 1.  
 DR PROSITE; PS00636; DnaJ\_1; 1.  
 DR PROSITE; PS00766; DnaJ\_2; 1.  
 SQ SEQUENCE 370 AA; 43260 MW; D5807ADA8078D95C CRC64;

Query Match 59.0%; Score 42.5; DB 5; Length 370;  
 Best Local Similarity 75.0%; Pred. No. 55;  
 Matches 9; Conservative 2; Mismatches 0; Indels 1; Gaps 1;

QY 1 RSNYDRSGW 12  
 DB 109 RSNYNR-FGDY 119

## RESULT 11

Q87UR0 PRELIMINARY; PRT; 37 AA.  
 AC Q87UR0; 01-JUN-2003 (TRENBLREL. 24, Created)  
 DT 01-JUN-2003 (TRENBLREL. 24, Last sequence update)  
 DT 01-JUN-2003 (TRENBLREL. 24, Last annotation update)  
 DE Hypothetical protein.  
 DE PSP05236.  
 GN Pseudomonas syringae (pv. tomato).  
 OS Pseudomonas syringae (pv. tomato).  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
 OC Pseudomonadaceae; Pseudomonas.  
 CX NCBI\_TaxID=323;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=DC3000.  
 RA Buell R., Joardar V., Khouri H., Fedorova N., Tran B., Russell D., RA Berry K., Utecherback T., Van Aken S., Feldblyum T., Gilm M., RA Dodson R., Deboy R., Durkin A., Kolonay J., Madupu R., Daugherty S., RA Brinkac L., Beanan M., Haft D., Selengut J., Nelson W., Davidson T., RA White O., Fraser C., Collier A.;  
 RT "Complete sequence of Pseudomonas syringae.";  
 RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AE016875; AA058662.1;  
 DR TIGR; PSP05236; -.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 37 AA; 4088 MW; F676CF84E728069 CRC64;

Query Match 58.3%; Score 42; DB 16; Length 37;  
 Best Local Similarity 72.7%; Pred. No. 5.4;  
 Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 RSNYDRSGW 12  
 DB 7 RSNPRLYGDY 17

RESULT 12  
 Q81QZ3 PRELIMINARY; PRT; 251 AA.



AC Q81Q23;  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE CG32496-PA.  
 GN CG32496.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidae; Drosophilidae; Drosophila.  
 OC NCBI\_Taxid=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams N.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang X., Chen L.X.,  
 RA Brandon R.C., Rogers Y.H., Blazek R.G., Champe M., Pfeiffer B.D.,  
 RA Man K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gaber G.L.,  
 RA Abril J.F., Aghayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,  
 RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshirei A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Munz D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,  
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson W., Skupski M.P., Smith T.,  
 RA Slier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Teocor C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.Y., Wasserman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodgett, Morley K.C., Wu D., Yang S., Yao Q.A., Ye J.,  
 RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of *Drosophila melanogaster*.";  
 RL Science 287:12185-2195(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Celniker S.E., Adams M.D., Krommiller B., Wan K.H., Holt R.A.,  
 RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,  
 RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,  
 RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,  
 RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Faifan D.,  
 RA Ferreira S., Frisoe E., Galie R.F., Garg N.S., George R.A.,  
 RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,  
 RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshirei A.,  
 RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,  
 RA Pacle J., Paragas V., Park S., Patel S., Pfeiffer B.,  
 RA Pounanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,  
 RA Stapleton M., Strong R., Svirskas R., Teocor C., Tyler D.,  
 RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;  
 RT "Sequencing of *Drosophila melanogaster* genome.";  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,  
 RA Hradecky P., Huang Y., Kaminker J.S., Prochnik J.S., Smith C.D.,

RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,  
 RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,  
 RA Krommiller B., Marshall B., Millburn G., Richter J., Russo S.,  
 RA Searle S.M.J., Smith E., Shu S., Smutnak F., Whitefield E.,  
 RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.,  
 RA "Annotation of *Drosophila melanogaster* genome.";  
 RT Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;  
 RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RA FlyBase;  
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
 DE EMBL; AE003507; AAN09447.1; -  
 DR FlyBase; FBgn052496; CG32496.  
 DR InterPro; IPR002181; Fibrinogen\_C.  
 DR Pfam; PF00147; fibrinogen\_C; 1.  
 DR SMART; SM00186; FBG; 1.  
 DR PROSITE; PS00514; FIBRIN\_AG\_C\_DOMAIN; 1.  
 DR PROSITE; PS00514; FIBRIN\_AG\_C\_DOMAIN; 1.  
 SQ SEQUENCE 251 AA; 29077 MW; FE8789C9D377BC39 CRC64;  
 Query Match 58.3%; Score 42; DB 5; Length 251;  
 Best Local Similarity 66.7%; Pred. No. 44;  
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 QY 4 NYDRSMDGY 12  
 DB 86 NFDKMDY 94  
 RESULT 13  
 Q7XOV1 PRELIMINARY; PRT; 356 AA.  
 AC Q7XOV1;  
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE OSUJB0079B02.1 protein.  
 DE OSUJB0079B02.1.  
 GN OSUJB0079B02.1.  
 OS Oryza sativa (Rice).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoideae; Oryzoae; Oryza.  
 OC NCBI\_TaxID=4530;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Han B., Feng Q., Huang Y.C., Li Y., Zhu J.J., Zhao Q., Hu X.,  
 RA Liu Y., Yu Z., Chen L., Fan D.L., Weng Q.J., Zhang L.,  
 RA Lu Y.Q., Yu S.L., Liu X.H., Lu T.T., Zhang Y.J., Lu Y., Li C., Li T.,  
 RA Zhang Y., Hu H., Jia P.X., Qian Y.M., Ying K., Zhou B., Wang Z.H.,  
 RA Hao P., Zhang L., Wu M., Zhang R.Q., Guan J.P., Fu G., Wang S.Y.,  
 RA Ren S.X., Cai Z., Lin W., Gu W.Q., Zhu G.F., Tu Y.F., Jia J., Yin H.F.,  
 RA Zhang Y., Li Z., Chen J., Kang H., Chen X.Y., Shao C.Y., Sun Y.,  
 RA Hu Q.P., Zhang X.L., Zhang W., Wang L.J., Ding C.W., Sheng H.H.,  
 RA Gu J.L., Chen S.T., Ni L., Zhu F.H., Hong G.F.,  
 RT Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AL606455; CAEO2969.1;  
 SQ SEQUENCE 356 AA; 39451 MW; 352A14B6B1F022C8 CRC64;  
 Query Match 58.3%; Score 42; DB 10; Length 356;  
 Best Local Similarity 50.0%; Pred. No. 64;  
 Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;  
 QY 1 RSNYDRSMDGY 12  
 DB 57 RSNYDRSMDGY 68  
 RESULT 14  
 Q9YX04 PRELIMINARY; PRT; 358 AA.  
 ID Q9YX04

AC Q9VX04; (1-MAY-2000 (TRENBLrel. 13, Created)  
 DT 01-OCT-2002 (TRENBLrel. 22, Last sequence update)  
 DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)  
 DE CG6788 protein (LP06940p).  
 GN CG6788.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 RX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Berkeley;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers J.H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abell J.F., Agbayani A., An H.-J., Andrews-Pianko C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.V., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borokova D., Botchan M.R., Bouck J., Brockstein P., Brodtner P.,  
 RA Butts K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Fabios B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K.J., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Dushin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Foster C., Garg R., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Mei M.-H., Ibegwam C.,  
 RA Jatala M., Kalusz F., Karpen G.H., Ke Z., Kennison V.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Laeko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusken D.R., Pacle J.M.,  
 RA Palazolo M., Pittman G.S., Pan S., Pollard V., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskaes R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zhang X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster";  
 RL Science 287:2185-2195(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC Celniker S.E., Adams M.D., Krommiller B., Wan K.H., Holt R.A.,  
 RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,  
 RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.V., Busam D.A.,  
 RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,  
 RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,  
 RA Ferrera S., Frise E., Galle R.F., Garg N.S., George R.A.,  
 RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howard T.J.,  
 RA Ibegwam C., Jatala M., Kruse D., Li P., Mattel B., Moshrefi A.,  
 RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,  
 RA Pacle J., Paragas V., Park S., Patel S., Pfeiffer B.,  
 RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,  
 RA Stapleton M., Strong R., Svirskaes R., Tector C., Tyler D.,  
 RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.,  
 RT "Sequencing of Drosophila melanogaster genome";  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Nipponbare;  
 RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,

RA Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,  
 RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,  
 RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,  
 RA Krommiller B., Marshall B., Milburn G., Richter J., Russo S.,  
 RA Searle S.M.J., Smith E., Shu S., Smurnick F., Whitfield E.,  
 RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.,  
 RT "Annotation of Drosophila melanogaster genome";  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter J.C.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RA FlyBase;  
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Berkeley;  
 RA Stapleton M., Brockstein P., Hong L., Agbayani A., Carlson J.,  
 RA Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,  
 RA George R., Gonzalez M., Guan H., Krommiller B., Li P., Liao G.,  
 RA Miranda A., Mungall C.J., Nunoo J., Pacle J., Paragas V., Park S.,  
 RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,  
 RA Celniker S.;  
 RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL, AE003507; AAF48780.2; -;  
 DR EMBL, A119053; AAM50913.1; -;  
 DR HSSP; P02671; 1F2D.  
 DR FlyBase; FBgn0030880; CG6788.  
 DR InterPro; IPR002181; Fibrinogen\_C.  
 DR Pfam; PF00147; fibrinogen\_C\_1.  
 DR SMART; SM00186; FBG; 1  
 DR PROSITE; PS00514; FIBRIN\_AG\_C\_DOMAIN; 1  
 DR SEQUENCE 358 AA; 41294 MW; 602452EAD9126670 CRC64;  
 SQ  
 Query Match 58.3%; Score 42; DB 5; Length 358;  
 Best local Similarity 66.7%; Pred. No. 64;  
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 DB 193 NFDGKMDY 201  
 QY 4 NYDRSGMDY 12  
 ID 193 NFDGKMDY 201  
 AC 08LNE3; PRELIMINARY; PRT; 395 AA.  
 DT 01-OCT-2002 (TRENBLrel. 22, Created)  
 DT 01-OCT-2002 (TRENBLrel. 22, Last sequence update)  
 DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)  
 DE Hypothetical protein.  
 GN OSJNB004410.6  
 OS Oryza sativa (Japonica cultivar-group).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Eriocaridaceae; Oryzaceae; Oryza.  
 OX NCBI\_TaxID=39947;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Nipponbare;  
 RA McCombie W.R., de la Bastide M., Spiegel L., Preston R., Ferraro K.,  
 RA Kull K., Nascimben L., Zuberlin T., Balla V., Bell M., Baker J.,  
 RA Miller B., Katzenberger F., Muller S., King L., Sullivan P., Yang C.,  
 RA Dike S., O'Shaughnessy A., Palmer L., Dedina N.,  
 RT "Genomic sequencing for Oryza sativa, Nipponbare strain, clone  
 OSJNB004410, from chromosome 10, complete sequence";  
 RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Nipponbare;  
 RA The Rice Chromosome 10 Sequencing Consortium;

RT "In-depth view of structure, activity, and evolution of rice  
RT chromosome 10."  
RI Science 300:1566-1569(2003).  
RN (3)  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Nipponbare;  
RA Buehl C.R., Ming R.A., McCombie W.R., Messing J., Yuan Q.,  
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AC083943; AAM74266.1; -  
DR EMBL; AB017091; AAP53659.1; -  
DR Gramene; Q81NE3; -  
KW Hypothetical protein.  
SQ SEQUENCE 395 AA; 43122 MW; 3C9BD1816E36E9F2 CRC64;  
Query Match 58.3%; Score 42; DB 10; Length 395;  
Best Local Similarity 50.0%; Pred. No. 71;  
Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;  
QY 1 RRSNYDRSWGDY 12  
| : | | | : | : |  
Db 56 RSKNFDPMWGEW 67

Search completed: October 5, 2004, 08:13:31  
Job time : 10.3519 secs

This Page Blank (uspio)

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2004, 07:04:10 : Search time 9.12963 seconds  
(without alignments)  
371.381 Million cell updates/sec

Title: US-09-805-290A-14  
Perfect score: 72  
Sequence: 1 RRSNYDRSWGDY 12

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_29Jan04:\*  
1: geneseqp1980s:\*  
2: geneseqp1990s:\*  
3: geneseqp2000s:\*  
4: geneseqp2001s:\*  
5: geneseqp2002s:\*  
6: geneseqp2003as:\*  
7: geneseqp2003bs:\*  
8: geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	72	100.0	12	AAE10547	Aae10547 Llana spe
2	72	100.0	129	AAE10551	Aae10551 HPL inhib
3	46	63.9	491	ABP65839	Abp65839 Bifidobac
4	44	61.1	12	AAE10548	Aae10548 Llana spe
5	44	61.1	129	AAE10558	Aae10558 HPL inhib
6	43	59.7	110	ABG01095	Abg01095 Novel hum
7	42	58.3	137	AAO05385	Aao05385 Human pol
8	42	58.3	274	ABB62687	Abb62687 Drosophila
9	42	58.3	318	AAW80057	Aaw80057 Human pro
10	42	58.3	3657	ABR82900	AbR82900 Human SMC
11	41	56.9	314	ABR65759	ABR65759 Rat hepat
12	41	56.9	415	ABW71400	ABW71400 Staphyloc
13	40	55.6	236	ABG23665	ABG23665 Novel hum
14	40	55.6	293	ABG15428	ABG15428 Novel hum
15	40	55.6	359	ABG02099	ABG02099 Novel hum
16	40	55.6	365	ABG23681	ABG23681 Novel hum
17	40	55.6	420	ABG15146	ABG15146 Novel hum
18	40	55.6	420	ABG15646	ABG15646 Novel hum
19	40	55.6	475	AAU36782	AAU36782 Staphyloc
20	40	55.6	480	ABU34289	ABU34289 Protein e
21	40	55.6	481	AAAG8152	AAAG8152 Mycobacte
22	40	55.6	481	ABU36657	ABU36657 Protein e
23	40	55.6	481	ABU35926	ABU35926 Protein e
24	40	55.6	481	ABR55186	ABR55186 Amino aci
25	40	55.6	481	ABW71857	ABW71857 Staphyloc

26	40	55.6	486	AAAG91246	AAg91246 C glutami
27	40	55.6	505	AAU60585	AAu60585 Propionib
28	40	55.6	505	ABM57104	ABm57104 Propionib
29	40	55.6	513	ABU25822	ABu25822 Protein e
30	40	55.6	529	ABG21716	ABg21716 Novel hum
31	40	55.6	531	AAAM25619	AAm25619 Human pro
32	40	55.6	535	ABU34517	ABu34517 Protein e
33	40	55.6	554	AAV53788	AAv53788 A maize p
34	40	55.6	554	ABU27693	ABu27693 Protein e
35	40	55.6	561	AAAM93771	AAm93771 Human pol
36	40	55.6	561	AAAB94525	AAb94525 Human pro
37	40	55.6	561	ABP64745	ABp64745 Human pro
38	40	55.6	662	ABG03325	ABg03325 Novel hum
39	40	55.6	674	ABG06139	ABg06139 Novel hum
40	40	55.6	676	ABG12518	ABg12518 Novel hum
41	40	55.6	730	ABG21383	ABg21383 Novel hum
42	40	55.6	902	ABG02304	ABg02304 Novel hum
43	40	55.6	996	ABG05431	ABg05431 Novel hum
44	40	55.6	996	ABG02173	ABg02173 Novel hum
45	40	55.6	1027	ABG09150	ABg09150 Novel hum

## ALIGNMENTS

RESULT 1  
ID AAE10547 standard; peptide, 12 AA.  
XX AAE10547;  
AC AAE10547;  
XX 10-DEC-2001 (first entry)  
DT 10-DEC-2001 (first entry)  
XX  
DE Llana species antibody VH CDR3 #16.  
XX  
XX Llana antibody; camelid; anorectic; heavy chain variable domain; VHH,  
KW human dietary enzyme inhibitor; medicament; human pancreatic lipase; HPL;  
KW food; human gastric lipase; HGL; cosmetic control; body weight;  
KW complementarity determining region 3; CDR3.  
XX  
XX Lama sp.  
OS  
XX  
XX EPI134231-A1.  
PN  
XX  
PD 19-SEP-2001.  
XX  
XX 20-FEB-2001; 2001EP-00200703.  
PF  
XX  
XX 14-MAR-2000; 2000EP-00200930.  
PR  
XX (UNITL ) UNILEVER NV.  
PA (UNITL ) UNILEVER PLC.  
XX  
XX Bezemer S, Van De Burg M, De Haard JW, Tareilus E;  
PI WPI; 2001-572718/65.  
XX  
XX New antibody or its fragments for inhibiting human dietary enzymes,  
PT useful for cosmetic control of body weight of human beings, comprises  
PT heavy chain variable domain derived from immunoglobulin naturally devoid  
PT of light chains.  
XX  
XX Claim 4; Page 29; 37pp; English.  
XX  
XX The patent discloses antibodies or their fragments comprising a heavy  
XX chain variable domain (VHH) derived from an immunoglobulin naturally  
XX devoid of light chain specific for inhibiting human dietary enzymes. The  
XX antibodies of the invention are useful for the preparation of medicaments  
XX or food for inhibiting the activity of one or more human dietary enzymes  
XX especially human pancreatic lipase (HPL) or human gastric lipase (HGL)  
XX which are useful for the cosmetic control of body weight of human beings.  
XX The present peptide sequence is a complementarity determining region 3  
XX (CDR3) of llama species (camelid) antibody VHH region

XX Sequence 12 AA; 100.0%; Score 72; DB 4; Length 12;  
SQ Best Local Similarity 100.0%; Pred. No. 2e-05;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 RRSNYDRSNGDY 12  
1 RRSNYDRSNGDY 12  
DB 1 RRSNYDRSNGDY 12

RESULT 2  
AAE10551  
ID AAE10551 standard; peptide; 129 AA.  
XX AAE10551;  
AC AAE10551;  
XX 10-DEC-2001 (first entry)  
XX HPL inhibiting VHH fragment, HPL #11 from llama species.  
XX Llama antibody; camelid; anorectic; heavy chain variable domain; VHH;  
XX human dietary enzyme inhibitor; medicament; human pancreatic lipase; HPL;  
XX food; human gastric lipase; HGL; cosmetic control; body weight.  
XX Lama SP.  
XX  
XX Key Location/Qualifiers  
XX Region 31..35  
XX /label= CDR1  
XX /note= "Complementarity determining region 1"  
XX 50..64  
XX /label= CDR2  
XX /note= "Complementarity determining region 2"  
XX 98..109  
XX /label= CDR3  
XX /note= "Complementarity determining region 3"  
XX  
XX EPI134231-A1.  
XX 19-SEP-2001.  
XX  
XX 20-FEB-2001; 2001EP-00200703.  
XX  
XX 14-MAR-2000; 2000EP-00200930.  
XX  
XX (UNIL ) UNILEVER NV.  
XX (UNIL ) UNILEVER PLC.  
XX  
XX Bezemer S, Van De Burg M, De Haard JWM, Tareilus E;  
XX WPI; 2001-572718/65.  
XX  
XX New antibody or its fragments for inhibiting human dietary enzymes,  
XX useful for cosmetic control of body weight of human beings, comprises  
XX heavy chain variable domain derived from immunoglobulin naturally devoid  
XX of light chains.  
XX  
XX Example 2; Page 9; 37pp; English.  
XX  
XX The patent discloses antibodies or their fragments comprising a heavy  
XX chain variable domain (VHH) derived from an immunoglobulin naturally  
XX devoid of light chains specific for inhibiting human dietary enzymes. The  
XX antibodies of the invention are useful for the preparation of medicaments  
XX or food for inhibiting the activity of one or more human dietary enzymes  
XX especially human pancreatic lipase (HPL) or human gastric lipase (HGL)  
XX which are useful for the cosmetic control of body weight of human beings.  
XX The present peptide sequence is HPL inhibiting VHH fragment, HPL #11 from  
XX llama (camelid) species  
XX  
XX Sequence 129 AA;

Query Match 100.0%; Score 72; DB 4; Length 129;  
Best Local Similarity 100.0%; Pred. No. 0.00026;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 RRSNYDRSNGDY 12  
98 RRSNYDRSNGDY 109  
DB 98 RRSNYDRSNGDY 109

RESULT 3  
ABP65839  
ID ABP65839 standard; protein; 491 AA.  
XX AC ABP65839;  
XX  
XX 19-NOV-2002 (first entry)  
XX  
XX Bifidobacterium longum NCC2705 ORF amino acid sequence SEQ ID NO:583.  
XX  
XX Bifidobacterium longum NCC2705; Bifidobacterium; bacterial;  
XX anti-diarrhetic; antibacterial; inhibitor of Salmonella; detection;  
XX identification; lactic acid bacterium; diarrhoea; pathogenic bacteria;  
XX rotavirus; food composition; pharmaceutical composition.  
XX  
XX Bifidobacterium longum.  
XX  
XX EPI227152-A1.  
XX  
XX 31-JUL-2002.  
XX  
XX 30-JUN-2001; 2001EP-00102050.  
XX  
XX 30-JUN-2001; 2001EP-00102050.  
XX  
XX (NEST ) SOC PROD NESTLE SA.  
XX  
XX WPI; 2002-668397/72.  
XX  
XX Novel polynucleotide comprising Bifidobacterium genome sequence useful as  
XX a probe or primer for detecting and/or identifying Bifidobacterium longum  
XX in a biological sample.  
XX  
XX Claim 3; SEQ ID NO 583; 80pp; English.  
XX  
XX The present invention describes a polynucleotide (I) comprising a  
XX sequence of a Bifidobacterium genome selected from the nucleotide  
XX sequences given in ABQ81842 and ABQ81843, or a sequence exhibiting at  
XX least 90% identity or which hybridises with the sequences given in  
XX ABQ81842 and ABQ81843. Also described is a polynucleotide (II) encoding a  
XX fusion protein, comprising a sequence selected from 1097 sequences given  
XX in ABP65258 to ABP6354, ligated in frame to a polynucleotide encoding a  
XX heterologous polypeptide. (I) has anti-diarrhetic and antibacterial  
XX activities, and can be used as an inhibitor of Salmonella. (I) (which is  
XX a probe) is useful for the detection and/or identification of  
XX Bifidobacterium longum in a biological sample. A carrier containing the  
XX lactic acid bacterium Bifidobacterium longum NCC2705 (CNCM I-2618) can be  
XX used for preventing and/or treating diarrhoea brought about by pathogenic  
XX bacteria and/or rotavirus. The carrier is a food composition selected  
XX from milk, yogurt, curd, cheese, fermented milks, milk based fermented  
XX products, ice-creams, fermented cereal based products, milk based  
XX powders, infant formula, pet food or a pharmaceutical composition  
XX selected from tablets, liquid bacterial suspensions, dried oral  
XX supplement, wet oral supplement, dry tube feeding or wet tube feeding.  
XX (I) is useful in DNA arrays or chips to carry out analysis of the  
XX expression of the Bifidobacterium gene. ABQ81844 to ABQ81850 represent  
XX Bifidobacterium related nucleotide sequences given in the sequence  
XX listing from the present invention but not mentioned further within the  
XX specification. N.B. The sequence data for this patent is not represented  
XX in the printed specification but is based on sequence information  
XX supplied by the European Patent Office  
XX  
XX Sequence 491 AA;

[illegible]

AC	AAE10558;
XX	
DT	10-DEC-2001 (first entry)
XX	
DE	HPL inhibiting VHH fragment, HPL #22 from llama species.
XX	
KW	Llama antibody; camelid; anorectic; heavy chain variable domain; VHH;
KM	human dietary enzyme inhibitor; medicament; human pancreatic lipase; HPL;
KW	food; human gastric lipase; HGL; cosmetic control; body weight.
XX	
OS	Lama sp.
XX	
XX	Location/Qualifiers
FH	Key
FT	Region
FT	/label= CDR1
FT	/note= "Complementarity determining region 1"
FT	Region
FT	/label= CDR2
FT	/note= "Complementarity determining region 2"
FT	Region
FT	/label= CDR3
FT	/note= "Complementarity determining region 3"
PN	EPIJ34231-A1.
PD	19-SEP-2001.
XX	
PF	20-FEB-2001; 2001EP-00200703.
PR	14-MAR-2000; 2000EP-00200930.
XX	
PA	(UNITL ) UNILEVER NV.
XX	(UNIL ) UNILEVER PLC.
P1	Bezemer S, Van De Burg M, De Haard JWM, Tareilus E;
PI	WPI; 2001-572718/65.
DR	
XX	New antibody or its fragments for inhibiting human dietary enzymes,
PT	useful for cosmetic control of body weight of human beings, comprises
PT	heavy chain variable domain derived from immunoglobulin naturally devoid
PT	of light chains.
PS	Example 2; Page 10; 37pp; English.
CC	The patent discloses antibodies or their fragments comprising a heavy
CC	chain variable domain (VHH) derived from an immunoglobulin naturally
CC	devoid of light chains specific for inhibiting human dietary enzymes. The
CC	antibodies of the invention are useful for the preparation of medicaments
CC	or food for inhibiting the activity of one or more human dietary enzymes
CC	especially human pancreatic lipase (HPL) or human gastric lipase (HGL)
CC	which are useful for the cosmetic control of body weight of human beings.
CC	The present peptide sequence is HPL inhibiting VHH fragment, HPL #22 from
CC	llama (camelid) species
XX	
SSQ	Sequence 129 AA:
Query Match	61.1%; Score 44; DB 4; Length 129;
Best Local Similarity	70.0%; Pred. No. 14;
Matches	7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY	3 SNYDSRGDY 12   :             :
DB	100 SSYDGSMDY 109
RESULT 6	
ABG01095	
ID	ABG01095 standard; Protein; 110 AA.
AC	ABG01095;
XX	
DT	13-FEB-2002 (first entry)

```

XX DE Novel human diagnostic protein #1086.
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX KW food supplement; medical imaging; diagnostic; genetic disorder.
XX OS Homo sapiens.
XX PN WO200175067-A2.
XX PD 11-OCT-2001.
XX PF 30-MAR-2001; 2001WO-US008631.
XX PR 31-MAR-2000; 2000US-00540217.
XX PR 23-AUG-2000; 2000US-00649167.
XX PA (HYSE-) HYSEQ INC.
XX PI Drmanac RT, Liu C, Tang YT;
XX DR WPI; 2001-639362/73.
XX DR N-PSDB; AAS65282.
XX PT New isolated polynucleotide and encoded polypeptides, useful in
XX PT diagnostics, forensics, gene mapping, identification of mutations
XX PT responsible for genetic disorders or other traits and to assess
XX PT biodiversity.
XX PS Claim 20; SEQ ID NO 31454; 103bp; English.
XX SS
XX CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
XX CC sequences. (I) is useful as hybridisation probes, polymerase chain
XX CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
XX CC and in recombinant production of (II). The polynucleotides are also used
XX CC in diagnostics as expressed sequence tags for identifying expressed
XX CC genes. (I) is useful in gene therapy techniques to restore normal
XX CC activity of (II) or to treat disease states involving (II). (II) is
XX CC useful for generating antibodies against it, detecting or quantitating a
XX CC polypeptide in tissue, as molecular weight markers and as a food
XX CC supplement. (II) and its binding partners are useful in medical imaging
XX CC of sites expressing (II). (I) and (II) are useful for treating disorders
XX CC involving aberrant protein expression or biological activity. The
XX CC polypeptide and polynucleotide sequences have applications in
XX CC diagnostics, forensics, gene mapping, identification of mutations
XX CC responsible for genetic disorders or other traits to assess biodiversity
XX CC and to produce other types of data and products dependent on DNA and
XX CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
XX CC amino acid sequences of the invention. Note: The sequence data for this
XX CC patent did not appear in the printed specification, but was obtained in
XX CC electronic format directly from WIPO at
XX CC ftp.wipo.int/pub/published_pct_sequences
XX SQ
XX Sequence 110 AA;
XX
XX Query Match 59.7%; Score 43; DB 4; Length 110;
XX Best Local Similarity 87.5%; Pred. No. 17;
XX Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
XX QY 3 SNYDRSWG 10
XX | | | | |
XX Db 99 SNYRSWG 106
XX
XX RESULT 7
XX ID AAO05385 standard; protein; 137 AA.
XX AC AAO05385;
XX DT 06-NOV-2001 (first entry)
XX XX Human polypeptide SEQ ID NO 19277.
XX DE

```

```

XX KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
XX KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
XX KW tissue growth factor; immunomodulatory; cancer; leukaemia;
XX KW nervous system disorders; arthritis; inflammation.
XX OS Homo sapiens.
XX PN WO200164835-A2.
XX PD 07-SEP-2001.
XX PF 26-FEB-2001; 2001WO-US004927.
XX PR 28-FEB-2000; 2000US-00515126.
XX PR 18-MAY-2000; 2000US-00577409.
XX PA (HYSE-) HYSEQ INC.
XX PI Tang YT, Liu C, Drmanac RT;
XX DR WPI; 2001-514838/56.
XX DR N-PSDB; AAI85316.
XX PT Isolated nucleic acids and polypeptides, useful for preventing diagnosing
XX PT and treating e.g. leukaemia, inflammation and immune disorders.
XX PS Claim 20; SEQ ID NO 19277; 1399bp + Sequence listing; English.
XX SS
XX CC The invention relates to human polynucleotides (AAI79941-AAI93641) and
XX CC the encoded proteins (AAO00010-AAO13910) that exhibit activity relating to
XX CC cytokine, cell proliferation or cell differentiation or which may induce
XX CC production of other cytokines in other cell populations. The
XX CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
XX CC peptide therapy. The polypeptides have various cytokine-like activities,
XX CC e.g. stem cell growth factor activity, haematopoiesis regulating
XX CC activity, tissue growth factor activity, immunomodulatory activity and
XX CC activity/inhibin activity and may be useful in the diagnosis and/or
XX CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
XX CC inflammation. Note: The sequence data for this patent did not form part
XX CC of the printed specification, but was obtained in electronic format
XX CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX SQ
XX Sequence 137 AA;
XX
XX Query Match 58.3%; Score 42; DB 4; Length 137;
XX Best Local Similarity 54.5%; Pred. No. 32;
XX Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
XX
XX QY 1 RSNYDRSWG 11
XX | | | | |
XX Db 103 KKNXDRSGMD 113
XX
XX RESULT 8
XX ID ABB62687 standard; protein; 274 AA.
XX AC ABB62687;
XX DT 26-MAR-2002 (first entry)
XX XX Drosophila melanogaster polypeptide SEQ ID NO 14853.
XX DE Drosophila melanogaster polypeptide SEQ ID NO 14853.
XX KW Drosophila; developmental biology; cell signalling; insecticide;
XX KW pharmaceutical.
XX OS Drosophila melanogaster.
XX PN WO200171042-A2.
XX PD 27-SEP-2001.
XX DE

```



PF 23-MAR-2001; 2001WO-US009231.  
 XX  
 PR 23-MAR-2000; 2000US-0191637P.  
 PR 11-JUL-2000; 2000US-00614150.  
 XX  
 PA (PEKE) PE CORP NY.  
 XX  
 PI Venter JC, Adams M, Li PWD, Myers EW;  
 XX WPI; 2001-656860/75.  
 DR N-PSDB; ABL06790.  
 XX  
 PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signaling and cell-cell  
 PT interactions.  
 XX  
 PS Disclosure; SEQ ID NO 14853; 21bp + Sequence Listing; English.  
 XX  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signaling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 CC sequences (ABL01840-ABL16175) and the encoded proteins (ABBS7737-  
 CC ABB72072). The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WIPO at fep.wipo.int/pub/published\_pct\_sequences  
 CC  
 SQ Sequence 274 AA;  
 XX  
 QY Query Match 58.3%; Score 42; DB 4; Length 274;  
 Best Local Similarity 66.7%; Pred. No. 68;  
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 DB 4 NYDRSWGDY 12  
 189 NPDGRGWKDY 197  
 XX  
 RESULT 9  
 AAM80067  
 ID AAM80067 standard; protein; 318 AA.  
 XX  
 AC AAM80067;  
 XX  
 DT 06-NOV-2001 (first entry)  
 XX  
 DE Human protein SEQ ID NO 3713.  
 XX  
 KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
 KW nervous system disorder; arthritis; inflammation.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200157190-A2.  
 XX  
 PD 09-AUG-2001.  
 XX  
 PF 05-FEB-2001; 2001WO-US004098.  
 XX  
 PR 03-FEB-2000; 2000US-00496914.  
 PR 27-APR-2000; 2000US-00560875.  
 PR 20-JUN-2000; 2000US-00598075.  
 PR 19-JUL-2000; 2000US-00620325.  
 PR 01-SEP-2000; 2000US-00654936.  
 PR 15-SEP-2000; 2000US-00663551.  
 PR 20-OCT-2000; 2000US-00693325.  
 PR 30-NOV-2000; 2000US-00728422.  
 XX  
 PA (HYSE-) HYSEQ INC.

XX  
 PT Tang YT, Liu C, Dermanac RT, Asundi V, Zhou P, Xu C, Cao Y;  
 PI Ma Y, Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;  
 PI Xue H, Yang Y, Wejthman T, Goodrich R;  
 XX  
 DR WPI; 2001-476283/51.  
 DR N-PSDB; AAK53200.  
 XX  
 PT Nucleic acids encoding polypeptides with cytokine-like activities, useful  
 PT in diagnosis and gene therapy.  
 XX  
 PS Claim 20; Page 416; 6221pp; English.  
 XX  
 CC The invention relates to polynucleotides (AAK51456-AAK53435) and the  
 CC encoded polypeptides (AAM78323-AAM80302) that exhibit activity relating to  
 CC cytokine, cell proliferation or cell differentiation or which may induce  
 CC production of other cytokines in other cell populations. The  
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
 CC peptide therapy. The polypeptides have various cytokine-like activities,  
 CC e.g. stem cell growth factor activity, haematopoiesis regulating  
 CC activity, tissue growth factor activity, immunomodulatory activity and  
 CC activin/inhibin activity and may be useful in the diagnosis and/or  
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
 CC inflammation. Notes: Records for SEQ ID NO 2110 (AAK52581), 2111  
 CC (AAK52582) and 3666 (AAM80020) are omitted as the relevant pages from the  
 CC sequence listing were missing at the time of publication  
 CC  
 SQ Sequence 318 AA;  
 XX  
 QY Query Match 58.3%; Score 42; DB 4; Length 318;  
 Best Local Similarity 60.0%; Pred. No. 80;  
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
 DB 3 SNYDRSWGDY 12  
 135 TNYSRSWNDW 144  
 XX  
 RESULT 10  
 ABB82900  
 ID ABB82900 standard; protein; 3657 AA.  
 XX  
 AC ABB82900;  
 XX  
 DT 31-MAR-2003 (first entry)  
 XX  
 DE Human SMG-1 protein.  
 XX  
 KW SMG-1; phosphatidylinositol kinase-associated kinase; cytostatic;  
 KW antibiotic; cancer; human; enzyme.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200295025-A1.  
 XX  
 PD 28-NOV-2002.  
 XX  
 PF 22-NOV-2001; 2001WO-JP010234.  
 XX  
 PR 24-MAY-2001; 2001JP-00156088.  
 XX  
 PA (NISC-) JAPAN SCI & TECHNOLOGY CORP.  
 XX  
 PI Ohno S;  
 XX  
 DR WPI; 2003-120862/11.  
 DR N-PSDB; ABB24084.  
 XX  
 PT New SMG-1 (phosphatidylinositol kinase-associated kinase), for  
 PT diagnosing and screening drugs for pathological conditions caused by  
 PT formation of an early transcription termination codon e.g. small-cell  
 PT lung cancer.  
 XX

PS Claim 1, Page 86-107, 125pp; Japanese.  
 XX  
 CC The invention relates to a novel SMG-1 (phosphatidylinositol kinase-  
 CC associated kinase) polypeptide and encoding polynucleotide. The SMG-1  
 CC protein and polynucleotide are useful in diagnosis of, and screening for  
 CC drugs for, pathological conditions caused by formation of an early  
 CC transcription termination codon due to a nonsense mutation, e.g. small-  
 CC cell lung cancer. The present sequence represents a human SMG-1 protein  
 XX  
 SQ Sequence 3657 AA;  
 Query Match 58.3%; Score 42; DB 6; Length 3657;  
 Best Local Similarity 60.0%; Pred. No. 1.1e+03;  
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
 OY 3 SNYDRSGWDY 12  
 DB 15 TNYRSRWMDW 24  
 RESULT 11  
 ID AAR65759 standard; protein; 314 AA.  
 XX  
 AC AAR65759;  
 XX  
 DT 25-MAR-2003 (revised)  
 DT 17-MAY-1995 (first entry)  
 XX  
 DE Rat hepatic parenchymal cell growth factor.  
 XX  
 KW Hepatic parenchymal cell growth factor; HPGF; liver diseases;  
 KW liver cancer; cirrhosis.  
 XX  
 OS Rattus rattus.  
 XX  
 FH Key Location/Qualifiers  
 FT Peptide 1..24  
 FT /label= sig\_peptide  
 XX  
 PN WO9421678-A1.  
 XX  
 PD 29-SEP-1994.  
 XX  
 PF 22-MAR-1994; 94WO-JP000455.  
 PF  
 PR 23-MAR-1993; 93JP-00063905.  
 PR  
 PA (TAIS ) TAISHO PHARM CO LTD.  
 PA  
 PI Hara H, Yoshimura H, Matsuki Y, Shindo S, Hanada X;  
 PI  
 DR WPI; 1994-316940/39.  
 DR N-PSDB; AAO77818.  
 DR  
 XX Hepatic parenchymal cell growth promoter peptide - is isolated from human  
 PT or animal liver cell or produced by recombinant techniques and used for  
 PT therapy of liver diseases.  
 XX  
 PS Claim 3, Page 34; 47pp; Japanese.  
 XX  
 CC AAO77818 encodes AAR65759 rat hepatic parenchymal cell growth factor  
 CC (HPGF). The protein or the N-terminal peptide (AAR65758) may be used in  
 CC the diagnosis and treatment of liver diseases, such as liver cancer and  
 CC cirrhosis. (Updated on 25-MAR-2003 to correct PN field.)  
 XX  
 SQ Sequence 314 AA;  
 Query Match 56.9%; Score 41; DB 2; Length 314;  
 Best Local Similarity 50.0%; Pred. No. 1.2e+02;  
 Matches 8; Conservative 2; Mismatches 2; Indels 4; Gaps 1;  
 OY 1 RRS----NYDRSGWDY 12

DB 127 RRDGSENFNRGMNDY 142  
 RESULT 12  
 ID AEM71400  
 ID AEM71400 standard; protein; 415 AA.  
 XX  
 AC AEM71400;  
 XX  
 DT 20-NOV-2003 (first entry)  
 DT  
 XX Staphylococcus aureus protein #640.  
 DE  
 XX Staphylococcus aureus protein #640.  
 XX  
 KW Antibacterial; vaccine; gene therapy; infection; sepsis; diagnosis;  
 KW enzymatic assay; antibiotic target.  
 XX  
 OS Staphylococcus aureus.  
 XX  
 PN WO200294868-A2.  
 PN  
 PD 28-NOV-2002.  
 PD  
 XX  
 PF 27-MAR-2002; 2002MO-IB002637.  
 PF  
 PR 27-MAR-2001; 2001GB-00007661.  
 PR  
 PA (CHIR-) CHIRON SPA.  
 PA  
 PI Masignani V, Mora M, Scarselli M;  
 PI  
 DR WPI; 2003-120786/11.  
 DR N-PSDB; ACF72960.  
 DR  
 XX New Staphylococcus aureus protein, useful as a vaccine for treating or  
 PT preventing Staphylococcal infection, specifically an infection caused by  
 PT S. aureus, e.g. sepsis.  
 PT  
 XX  
 PS Claim 1, SEQ ID NO 1280; 49pp; English.  
 XX  
 CC The invention relates to novel genes and encoded proteins from  
 CC Staphylococcus aureus. A composition comprising the S. aureus protein, a  
 CC nucleic acid encoding the protein, or an antibody to the protein, is  
 CC useful as a pharmaceutical, particularly as a vaccine for treating or  
 CC preventing infection due to Staphylococcus bacteria, specifically an  
 CC infection caused by S. aureus. The composition is particularly useful for  
 CC treating or preventing sepsis in a patient. The composition can also be  
 CC used for diagnostics. The protein is also used in an assay for enzymatic  
 CC studies and as a target for antibiotics. This sequence represents one of  
 CC the novel S. aureus proteins of the invention  
 XX  
 SQ Sequence 415 AA;  
 Query Match 56.9%; Score 41; DB 6; Length 415;  
 Best Local Similarity 63.6%; Pred. No. 1.6e+02;  
 Matches 7; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 OY 2 RSNYDRSGWDY 12  
 DB 363 RSOYQASWTDY 373  
 RESULT 13  
 ID ABG23665  
 ID ABG23665 standard; protein; 236 AA.  
 XX  
 AC ABG23665;  
 XX  
 DT 18-FEB-2002 (first entry)  
 DT  
 XX Novel human diagnostic protein #23656.  
 DE  
 XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW

KM food supplement; medical imaging; diagnostic; genetic disorder.  
 XX Homo sapiens.  
 OS WO200175067-A2.  
 XX  
 XX  
 PD 11-OCT-2001.  
 XX  
 XX 30-MAR-2001; 2001WO-US008631.  
 PF 31-MAR-2000; 2000US-00540217.  
 PR 23-AUG-2000; 2000US-00649167.  
 XX  
 XX (HYSE-) HYSEQ INC.  
 PA  
 PI Drmanac RT, Liu C, Tang YT.  
 XX  
 XX WPI; 2001-639362/73.  
 DR N-PSDB; AAS87852.  
 XX  
 XX  
 PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity.  
 PS  
 PS Claim 20; SEQ ID NO 54024; 103bp; English.  
 XX  
 XX The invention relates to isolated polynucleotide (I) and polypeptide (II)  
 CC sequences. (I) is useful as hybridisation probes, polymerase chain  
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
 CC and in recombinant production of (II). The polynucleotides are also used  
 CC in diagnostics as expressed sequence tags for identifying expressed  
 CC genes. (I) is useful in gene therapy techniques to restore normal  
 CC activity of (II) or to treat disease states involving (II). (II) is  
 CC useful for generating antibodies against it, detecting or quantitating a  
 CC polypeptide in tissue, as molecular weight markers and as a food  
 CC supplement. (II) and its binding partners are useful in medical imaging  
 CC of sites expressing (II). (I) and (II) are useful for treating disorders  
 CC involving aberrant protein expression or biological activity. The  
 CC polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic  
 CC amino acid sequences of the invention. Note: The sequence data for this  
 CC patent did not appear in the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 236 AA;  
 QY  
 DB 1 RRSNYDRSMGD 11  
 48 RRSNYSELMED 58  
 55.6%; Score 40; DB 4; Length 236;  
 Best Local Similarity 63.6%; Pred. No. 1.3e+02;  
 Matches 7; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 RESULT 14  
 ID ABG15428 standard; protein; 293 AA.  
 XX  
 AC ABG15428;  
 XX  
 DT 18-FEB-2002 (first entry)  
 XX  
 DE Novel human diagnostic protein #15419.  
 XX  
 KM Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KM food supplement; medical imaging; diagnostic; genetic disorder.  
 XX

OS Homo sapiens.  
 XX  
 XX WO200175067-A2.  
 XX  
 XX  
 PD 11-OCT-2001.  
 XX  
 XX 30-MAR-2001; 2001WO-US008631.  
 PF 31-MAR-2000; 2000US-00540217.  
 PR 23-AUG-2000; 2000US-00649167.  
 XX  
 XX (HYSE-) HYSEQ INC.  
 PA  
 PI Drmanac RT, Liu C, Tang YT.  
 XX  
 XX WPI; 2001-639362/73.  
 DR N-PSDB; AAS79615.  
 XX  
 XX  
 PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity.  
 PS  
 PS Claim 20; SEQ ID NO 45787; 103bp; English.  
 XX  
 XX The invention relates to isolated polynucleotide (I) and polypeptide (II)  
 CC sequences. (I) is useful as hybridisation probes, polymerase chain  
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
 CC and in recombinant production of (II). The polynucleotides are also used  
 CC in diagnostics as expressed sequence tags for identifying expressed  
 CC genes. (I) is useful in gene therapy techniques to restore normal  
 CC activity of (II) or to treat disease states involving (II). (II) is  
 CC useful for generating antibodies against it, detecting or quantitating a  
 CC polypeptide in tissue, as molecular weight markers and as a food  
 CC supplement. (II) and its binding partners are useful in medical imaging  
 CC of sites expressing (II). (I) and (II) are useful for treating disorders  
 CC involving aberrant protein expression or biological activity. The  
 CC polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic  
 CC amino acid sequences of the invention. Note: The sequence data for this  
 CC patent did not appear in the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 293 AA;  
 QY  
 DB 1 RRSNYDRSMGD 11  
 48 RRSNYSELMED 58  
 55.6%; Score 40; DB 4; Length 293;  
 Best Local Similarity 63.6%; Pred. No. 1.6e+02;  
 Matches 7; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 RESULT 15  
 ID ABG02099 standard; protein; 359 AA.  
 XX  
 AC ABG02099;  
 XX  
 DT 13-FEB-2002 (first entry)  
 XX  
 DE Novel human diagnostic protein #2090.  
 XX  
 KM Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KM food supplement; medical imaging; diagnostic; genetic disorder.  
 OS Homo sapiens.  
 XX

PN WO200175067-A2.  
 XX  
 PD 11-OCT-2001.  
 XX  
 PF 30-MAR-2001; 2001WO-US008631.  
 XX  
 PR 31-MAR-2000; 2000US-00540217.  
 XX PR 23-AUG-2000; 2000US-00649167.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 PI Dymnac RT, Liu C, Tang YT;  
 XX  
 DR WPI; 2001-639362/73.  
 DR N-PSDB; AAS66286.  
 XX  
 PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity.  
 XX  
 PS Claim 20; SEQ ID NO 32458; 103pp; English.  
 XX  
 CC The invention relates to isolated polynucleotide (I) and polypeptide (II)  
 CC sequences. (I) is useful as hybridization probes, polymerase chain  
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
 CC and in recombinant production of (II). The polynucleotides are also used  
 CC in diagnostics as expressed sequence tags for identifying expressed  
 CC genes. (I) is useful in gene therapy techniques to restore normal  
 CC activity of (II) or to treat disease states involving (II). (II) is  
 CC useful for generating antibodies against it, detecting or quantitating a  
 CC polypeptide in tissue, as molecular weight markers and as a food  
 CC supplement. (II) and its binding partners are useful in medical imaging  
 CC of sites expressing (II). (I) and (II) are useful for treating disorders  
 CC involving aberrant protein expression or biological activity. The  
 CC polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic  
 CC amino acid sequences of the invention. Note: The sequence data for this  
 CC patent did not appear in the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pcr\_sequences  
 XX  
 SQ Sequence 359 AA;

Query Match 55.6%; Score 40; DB 4; Length 359;  
 Best Local Similarity 63.6%; Pred. No. 2e+02;  
 Matches 7; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 OY 1 RRSNYDRSGD 11  
 |||||  
 DB 95 RRSNYSELMED 105

Search completed: October 5, 2004, 07:59:56  
 Job time : 13.1296 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2004, 07:36:16 : Search time 1.94444 Seconds  
(without alignments)  
593.639 Million cell updates/sec

Title: US-09-805-290A-15

Perfect score: 69

Sequence: 1 LISSYDGSWMDY 12

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	49	71.0	293	2 B70455	thiosulfate sulfur
2	47	68.1	277	2 G70809	probable thiosulfate
3	47	68.1	277	2 A87184	thiosulfate sulfur
4	45	65.2	466	2 T04976	hypothetical prote
5	45	65.2	1415	2 T21244	Zyg-9 protein - Ca
6	44	63.8	278	2 C90253	thiosulfate sulfur
7	43	62.3	300	2 F70196	flagellar hook-bas
8	42	60.9	276	2 G82436	probable thiosulfate
9	42	60.9	319	2 A70340	conserved hypothet
10	42	60.9	355	2 T14012	NADH2 dehydrogenas
11	42	60.9	355	2 T14014	NADH2 dehydrogenas
12	42	60.9	355	2 T14016	NADH2 dehydrogenas
13	42	60.9	2314	2 T28698	hypothetical prote
14	41	59.4	218	2 D72494	probable thiosulfate
15	41	59.4	284	2 AH0146	probable sulfurtra
16	41	59.4	286	2 D75547	thiosulfate sulfur
17	41	59.4	289	1 A37209	thiosulfate sulfur
18	41	59.4	295	2 S15081	thiosulfate sulfur
19	41	59.4	296	2 A87000	probable thiosulfate
20	41	59.4	297	2 JC4398	thiosulfate sulfur
21	41	59.4	297	2 JC5286	thiosulfate sulfur
22	41	59.4	297	2 D70980	probable sea prot
23	41	59.4	355	2 T12104	NADH2 dehydrogenas
24	41	59.4	355	2 T12100	NADH2 dehydrogenas
25	41	59.4	355	2 T13830	NADH2 dehydrogenas
26	41	59.4	355	2 T13831	NADH2 dehydrogenas
27	41	59.4	355	2 T11601	NADH2 dehydrogenas
28	41	59.4	355	2 T13833	NADH2 dehydrogenas
29	41	59.4	355	2 T13832	NADH2 dehydrogenas

30	41	59.4	355	2 T13904	NADH2 dehydrogenas
31	41	59.4	355	2 T13861	NADH2 dehydrogenas
32	41	59.4	355	2 T12114	NADH2 dehydrogenas
33	41	59.4	355	2 T12111	NADH2 dehydrogenas
34	41	59.4	355	2 T12118	NADH2 dehydrogenas
35	41	59.4	355	2 T12119	NADH2 dehydrogenas
36	41	59.4	355	2 T12112	NADH2 dehydrogenas
37	41	59.4	355	2 T12109	NADH2 dehydrogenas
38	41	59.4	355	2 T12107	NADH2 dehydrogenas
39	41	59.4	355	2 T13884	NADH2 dehydrogenas
40	41	59.4	355	2 T13988	NADH2 dehydrogenas
41	41	59.4	355	2 T11752	NADH2 dehydrogenas
42	41	59.4	355	2 T11762	NADH2 dehydrogenas
43	41	59.4	355	2 T11760	NADH2 dehydrogenas
44	41	59.4	355	2 T13989	NADH2 dehydrogenas
45	41	59.4	355	2 T14009	NADH2 dehydrogenas

#### ALIGNMENTS

##### RESULT 1

B70455 thiosulfate sulfurtransferase - Aquifex aeolicus

C/Species: Aquifex aeolicus

C/Date: 08-May-1998 #sequence\_revision 08-May-1998 #text\_change 21-Jan-2000

C/Accession: B70455

R/Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Oye

Nature 392, 353-358, 1998

A/Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.

A/Reference number: A70300; MUID:98196666; PMID:9537320

A/Accession: B70455

A/Status: preliminary; nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-293 <ADP>

A/Cross-references: GB:AE000757; NID:92984092; PIDN:AA07633.1; PID:92984095; GB:AE000657

A/Experimental source: strain VFS

C/Genetics:

A/Gene: rhdA2

C/Superfamily: thiosulfate sulfurtransferase

Query Match

Best Local Similarity 71.0%; Score 49; DB 2; Length 293;

Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 5 YDGSWMDY 12

DB 277 YDGSWMDY 284

##### RESULT 2

probable thiosulfate sulfurtransferase - Mycobacterium tuberculosis (strain H37RV)

C/Species: Mycobacterium tuberculosis

C/Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 20-Jun-2000

C/Accession: G70809; D70921

R/Colo, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.

; Connor, R.; Davies, R.; Devlin, K.; Felwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;

Rajandrem, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A/Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrett, B.G.

A/Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A/Reference number: A70500; MUID:98295987; PMID:9634230

A/Accession: G70809

A/Status: preliminary; nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-277 <COI>

A/Cross-references: GB:AL022004; GB:AL123456; NID:93261550; PIDN:CAA17621.1; PID:92916873

A/Experimental source: strain H37RV

A/Accession: D70921

A/Status: preliminary; nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A:Residues: 1-277 <CO2>  
A:Cross-references: GB:Z95150; GB:AL123456; NID:G3250708; PIDN:CA808374.1; PID:G2076692  
A:Experimental source: strain H37Rv  
C:Comment: There are two identical copies of this protein in the Mycobacterium tuberculosis  
C:Genetics:  
A:Gene: cysA2, cysA3  
C:Superfamily: thiosulfate sulfurtransferase

Query Match 68.1%; Score 47; DB 2; Length 277;  
Best Local Similarity 54.5%; Pred. No. 2.9;  
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 2 ISSYDGSNDY 12  
: : : : :  
DB 255 VKNDYGSMTXY 265

RESULT 3  
thiosulfate sulfurtransferase [imported] - Mycobacterium leprae  
C:Species: Mycobacterium leprae  
C:Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 17-May-2002  
C:Accession: A87184  
R: Cole, S.T.; Eigmeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; Ho  
R.; Davies, R.M.; Devlin, K.; Duthoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holroyd,  
eam, M.A.; Rutherford, K.M.  
Nature 409, 1007-1011, 2001  
A:Authors: Ruter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; Sq  
A:Title: Massive gene decay in the leprosy bacillus.  
A:Reference number: A86909; PMID:21128732; PMID:11234002  
A:Accession: A87184  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-277 <STO>  
A:Cross-references: GB:AL450380; NID:G13093731; PIDN:CAC31153.1; GSPDB:GN00147  
A:Genetics:  
A:Gene: cysA3  
C:Superfamily: thiosulfate sulfurtransferase

Query Match 68.1%; Score 47; DB 2; Length 277;  
Best Local Similarity 54.5%; Pred. No. 2.9;  
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 2 ISSYDGSNDY 12  
: : : : :  
DB 255 VKNDYGSMTXY 265

RESULT 4  
hypothetical protein T16L1.80 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 11-Jun-1999  
C:Accession: T04976  
R: Bevan, M.; Obermaier, B.; Deutschenbauer, S.; Piravandi, E.; Hohenisel, J.; Mewes, H.W.;  
submitted to the Protein Sequence Database, November 1998  
A:Reference number: Z15393  
A:Accession: T04976  
A:Molecule type: DNA  
A:Residues: 1-466 <BEV>  
A:Cross-references: EMBL:AD031394  
A:Experimental source: cultivar Columbia, BAC clone T16L1  
C:Genetics:  
A:Map position: 4  
A:Note: T16L1.80

Query Match 65.2%; Score 45; DB 2; Length 466;  
Best Local Similarity 54.5%; Pred. No. 11;  
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 ISSYDGSNDY 12  
: : : : :  
DB 134 VDEHDSMNSY 144

RESULT 5  
Zyg-9 protein - Caenorhabditis elegans  
T21244  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 31-Jan-2000  
C:Accession: T21244; T42392  
R: Sims, M.  
submitted to the EMBL Data Library, July 1995  
A:Reference number: Z19395  
A:Accession: T21244  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1415 <WIL>  
A:Cross-references: EMBL:Z50044; PIDN:CAA90359.1; GSPDB:GN00020; CESP:F22B5.7  
A:Experimental source: clone F22B5  
R: Matthews, L.; Baker, K.; Thierry-Mieg, D.; Kempnes, K.  
submitted to the EMBL Data Library, November 1997  
A:Reference number: Z22151  
A:Accession: T42392  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-1415 <MAT>  
A:Cross-references: EMBL:AF035197; PIDN:AA017865.1  
C:Genetics:  
A:Gene: zyg-9; F22B5.7  
A:Map position: 2  
A:introns: 64/3; 350/3; 543/3; 572/3; 623/3; 775/1; 890/3; 1061/3; 1372/2

Query Match 65.2%; Score 45; DB 2; Length 1415;  
Best Local Similarity 66.7%; Pred. No. 36;  
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 LISSYDGSNDY 12  
: : : : :  
DB 1291 LIASDPSMNEY 1302

RESULT 6  
C90253  
thiosulfate sulfurtransferase (cysA-1) [imported] - Sulfolobus solfataricus  
C:Species: Sulfolobus solfataricus  
C:Date: 24-May-2001 #sequence\_revision 24-May-2001 #text\_change 15-Jun-2001  
C:Accession: C90253  
R: She, O.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Aways, M.J.; Chan-  
Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, P.  
arret, R.A.; Kagan, M.A.; Jensen, C.W.; Van der Oost, J.  
submitted to GenBank, April 2001  
A:Description: Sulfolobus solfataricus complete genome.  
A:Reference number: A99139  
A:Accession: C90253  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-278 <KUR>  
A:Cross-references: GB:AB066441; NID:G13814200; PIDN:AAK41282.1; GSPDB:GN00155  
C:Genetics:  
A:Gene: cysA-1  
C:Superfamily: thiosulfate sulfurtransferase

Query Match 63.8%; Score 44; DB 2; Length 278;  
Best Local Similarity 58.3%; Pred. No. 8.9;  
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 LISSYDGSNDY 12  
: : : : :  
DB 251 LVRLYDSMVEY 262

RESULT 7  
F70196  
flagellar hook-basal body complex protein (FlhB) homolog - Lyme disease spirochete  
C:Species: Borrelia burgdorferi (Lyme disease spirochete)

C>Date: 13-Feb-1998 #sequence\_revision 13-Feb-1998 #text\_change 24-Nov-1999  
C/Accession: F70196  
R/Frazer, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White  
son, D.; Peterson, J.; Kervavagge, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt,  
Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horek, K.; Roberts, K.; Hatch, B.  
Nature 390, 580-586, 1997  
A/Authors: Smith, H.O.; Venter, J.C.  
A/Title: Genomic sequence of a Lyme disease spirochete, *Borrelia burgdorferi*.  
A/Reference number: A70100; MUID:98055943; PMID:9403685  
A/Accession: F70196  
A/Status: preliminary; nucleic acid sequence not shown; translation not shown  
A/Molecule type: DNA  
A/Residues: 1-300 <RLE>  
A/Cross-references: GB:AE001177; GB:AE000783; NID:g2688711; PIDN:AA67131.1; PID:g268872  
A/Experimental source: strain B31  
C/Superfamily: rod protein flagP

Query Match 62.3%; Score 43; DB 2; Length 300;  
Best Local Similarity 50.0%; Pred. No. 14;  
Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 LISSYDGSWMDY 12  
DB 194 LVSEYENSWENY 205

RESULT 8  
G82436  
probable thiosulfate sulfotransferase Ssea VCA0620 [imported] - *Vibrio cholerae* (strain  
C/Species: *Vibrio cholerae*  
C/Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Feb-2001  
C/Accession: G82436  
R/Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R.J.;  
chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragol, I.; Sellers, F.  
1. R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.  
Nature 406, 477-483, 2000  
A/Title: DNA Sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*.  
A/Reference number: A82035; MUID:20406833; PMID:10952301  
A/Accession: G82436  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-276 <HEI>  
A/Cross-references: GB:AE004392; GB:AE003853; NID:g9658031; PIDN:AAf6521.1; GSPDB:GN001  
A/Experimental source: serogroup O1; strain NA6961; biotype El Tor  
C/Genetics:  
A/Gene: VCA0620  
A/Map position: 2  
C/Superfamily: thiosulfate sulfotransferase

Query Match 60.9%; Score 42; DB 2; Length 276;  
Best Local Similarity 54.5%; Pred. No. 19;  
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 2 ISSYDGSWMDY 12  
DB 257 LSVYDGSWMDY 267

RESULT 9  
A70340  
conserved hypothetical protein aq\_438 - *Aquifex aeolicus*  
C/Species: *Aquifex aeolicus*  
C/Date: 08-May-1998 #sequence\_revision 08-May-1998 #text\_change 05-Nov-1999  
C/Accession: A70340  
R/Beckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Ov  
V.  
Nature 392, 353-358, 1998  
A/Title: The complete genome of the hyperthermophilic bacterium *Aquifex aeolicus*.  
A/Reference number: A70300; MUID:98196666; PMID:9537320  
A/Accession: A70340  
A/Status: preliminary; nucleic acid sequence not shown; translation not shown  
A/Molecule type: DNA  
A/Residues: 1-319 <AOF>

A/Cross-references: GB:AE000690; NID:g2983100; PIDN:AA06720.1; PID:g2983115; GB:AE000657  
A/Experimental source: strain VFS  
C/Genetics:  
A/Gene: aq\_438  
C/Superfamily: CBS homology  
F.203-250/Damain: CBS homology <CBS>

Query Match 60.9%; Score 42; DB 2; Length 319;  
Best Local Similarity 85.7%; Pred. No. 22;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 6 DGSWMDY 12  
DB 52 DGKXNDY 58

RESULT 10  
T14012  
NADH2 dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 5 - Damaster fortunei mitochondrion  
C/Species: mitochondrion Damaster fortunei  
C/Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 03-Jun-2002  
C/Accession: T14012  
R/Su, Z.H.; Tomioka, O.; Okamoto, M.; Osawa, S.  
Mol. Biol. Evol. 15, 1026-1039, 1998  
A/Title: Origin and diversification of hindwingsless Damaster ground beetles within the J  
A/Reference number: Z17319; MUID:98384842; PMID:9718730  
A/Accession: T14012  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 1-355 <SUZ>  
A/Cross-references: EMBL:D50426; PIDN:BA08922.1  
A/Experimental source: adult; thorax muscle  
C/Genetics:  
A/Genome: mitochondrion  
C/Superfamily: NADH dehydrogenase (ubiquinone) chain 5  
C/Keywords: membrane-associated complex; mitochondrion; NAD; oxidative phosphorylation; c

Query Match 60.9%; Score 42; DB 2; Length 355;  
Best Local Similarity 50.0%; Pred. No. 25;  
Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 LISSYDGSWMDY 12  
DB 299 LFTSPDGSWMDY 310

RESULT 11  
T14014  
NADH2 dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 5 - Damaster oxuroides mitochondrion  
C/Species: mitochondrion Damaster oxuroides  
A/Variety: isolate Gifu  
C/Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 03-Jun-2002  
C/Accession: T14014  
R/Su, Z.H.; Tomioka, O.; Okamoto, M.; Osawa, S.  
Mol. Biol. Evol. 15, 1026-1039, 1998  
A/Title: Origin and diversification of hindwingsless Damaster ground beetles within the J  
A/Reference number: Z17319; MUID:98384842; PMID:9718730  
A/Accession: T14014  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 1-355 <SUZ>  
A/Cross-references: EMBL:D50427; PIDN:BA08923.1  
A/Experimental source: isolate Gifu; adult; thorax muscle  
C/Genetics:  
A/Genome: mitochondrion  
C/Superfamily: NADH dehydrogenase (ubiquinone) chain 5  
C/Keywords: membrane-associated complex; mitochondrion; NAD; oxidative phosphorylation; c

Query Match 60.9%; Score 42; DB 2; Length 355;  
Best Local Similarity 50.0%; Pred. No. 25;  
Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 LISSYDGSWMDY 12

Db 299 LFTSFDOGMNEY 310

## RESULT 12

TI4016  
NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 5 - Damaster viridipennis mitochondrion  
C/Species: mitochondrion Damaster viridipennis  
C/Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 03-Jun-2002  
C/Accession: T14016  
R/Stu. Z.H.; Tomlinaga, O.; Okamoto, M.; Osawa, S.  
Mol. Biol. Evol. 15, 1026-1039, 1998  
A/Title: Origin and diversification of hindwingsless Damaster ground beetles within the U  
A/Reference number: 217319; MUID:99384842; PMID:9718730  
A/Accession: T14016  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 1355 <SUZ>  
A/Cross-references: EMBL:D50428; PIDN:BA008924.1  
A/Experimental source: adult; thorax muscle  
C/Genetics:  
A/Genome: mitochondrion  
C/Superfamily: NADH dehydrogenase (ubiquinone) chain 5  
C/Keywords: membrane-associated complex; mitochondrion; NAD; oxidative phosphorylation;

Query Match 60.9%; Score 42; DB 2; Length 355;  
Best Local Similarity 50.0%; Pred. No. 25;  
Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 LISSYDGSWMDY 12  
DB 299 LFTSFDOGMNEY 310

## RESULT 13

T28698  
hypothetical protein - Streptomyces coelicolor  
C/Species: Streptomyces coelicolor  
C/Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C/Accession: T28698  
R/Parkhill, J.; Bentley, S.D.; Barrall, B.G.; Rajandream, M.A.  
submitted to the EMBL Data Library, January 1999  
A/Reference number: 220512  
A/Accession: T28698  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 1-2314 <PAR>  
A/Cross-references: EMBL:AL023496; NID:e1292348; PID:e1292365; PIDN:CAA18915.1

Query Match 60.9%; Score 42; DB 2; Length 2314;  
Best Local Similarity 66.7%; Pred. No. 1.9e+02;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 4 SYDGSWMDY 12  
DB 1340 TYDGAWLDY 1348

## RESULT 14

D72434  
probable thiosulfate sulfotransferase APE2595 - Aeropyrum pernix (strain K1)  
C/Species: Aeropyrum pernix  
C/Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Jun-2000  
C/Accession: D72434  
R/Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takah  
awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; K  
DNA-Res. 6, 83-101, 1999  
A/Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyr  
A/Reference number: A72450; MUID:99310339; PMID:10382966  
A/Accession: D72434  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-218 <KAW>

A/Cross-references: DBJ:AP000064; NID:g5105945; PIDN:BA01612.1; PID:g5106301  
A/Experimental source: strain K1  
C/Genetics:  
A/Genome: APE2595  
C/Superfamily: thiosulfate sulfotransferase

Query Match 59.4%; Score 41; DB 2; Length 218;  
Best Local Similarity 45.5%; Pred. No. 21;  
Matches 5; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 2 ISSYDGSWMDY 12  
DB 194 VAVYDGSWSEW 204

## RESULT 15

AH0146  
probable sulfotransferase YP01194 [imported] - Yersinia pestis (strain CO92)  
C/Species: Yersinia pestis  
C/Date: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 09-Nov-2001  
C/Accession: AH0146  
R/Parkhill, J.; Wren, B.W.; Thomson, N.R.; Tlball, R.W.; Holden, M.T.G.; Prentice, M.B.;  
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.; F  
11, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrall, E  
Nature 413, 523-527, 2001  
A/Title: Genome sequence of Yersinia pestis, the causative agent of plague.  
A/Reference number: ABC001; MUID:21470413; PMID:11586360  
A/Accession: AH0146  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-284 <KUR>  
A/Cross-references: GB:AL590842; PIDN:CAC90035.1; PID:g15979257; GSPDB:GN00175  
C/Genetics:  
A/Genome: YP01194  
C/Superfamily: thiosulfate sulfotransferase

Query Match 59.4%; Score 41; DB 2; Length 284;  
Best Local Similarity 54.5%; Pred. No. 28;  
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 2 ISSYDGSWMDY 12  
DB 260 VSLYDGSWAEW 270

Search completed: October 5, 2004, 08:17:06  
Job time: 2.94444 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd

OM protein - protein search, using sw model

```
Run on:      October 5, 2004, 07:29:35 ; Search time 1.07407 Seconds
              (without alignments)
              581.749 Million cell updates/sec
```

Title: US-09-805-290A-13  
Perfect score: 69  
Sequence: 1 LISSYDGSWNDY 12

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

```
Searched:      141681 segs, 52070155 residues
Total number of hits satisfying chosen parameters: 141681
```

```

Minimum DB seq length: 0
Maximum DB seq length: 20000000000
Post-processing: Minimum Match 0%
                  Maximum Match 100%

```

Database : SwissProt\_42:\*

pred. 'No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Query No.	Score	Match	Length	DB	ID	Description
1	47	68.1	277	1	TH1_MYCBO	P59959 mycobacter
2	47	68.1	277	1	TH1_MYCLE	Q50036 mycobacter
3	47	68.1	277	1	TH1_MYCU	Q50793 mycobacter
4	47	68.1	281	1	TH1_SACR	O57573 mycobacter
5	41	59.4	289	1	TH1R_CHICK	P16355 saccharopo
6	41	59.4	295	1	TH1R_RAT	P55324 gallus gori
7	41	59.4	296	1	TH12_MYCLE	P46700 mycobacter
8	41	59.4	296	1	TH1R_CRICR	P46635 cricetusus
9	41	59.4	296	1	TH1R_HUMAN	Q16762 homo sapien
10	41	59.4	296	1	TH1R_MOUSE	P52186 mus musculi
11	41	59.4	297	1	TH12_MYCU	P56888 mycobacter
12	41	59.4	297	1	Y4MH_RHNSN	P55567 rhizobium
13	41	59.4	878	1	IL13_MOUSE	P26954 mus musculi
14	41	59.4	1046	1	CH1D_VIBRU	P26156 vibrio furn
15	40	58.0	601	1	YAN4_SCHPO	Q10070 schizosacc
16	39	56.5	290	1	Y489_CHLEP	Q52862 chlamydia
17	39	56.5	301	1	TH1R_CORFL	P71121 corynebact
18	39	56.5	320	1	TH1R_SYNP7	P27477 synechococ
19	39	56.5	588	1	YA71_SCHRO	O09758 schizosacc
20	39	56.5	896	1	CR1B_MOUSE	P26955 mus musculi
21	38	55.1	142	1	IL3_CALUA	Q28334 callithrix
22	38	55.1	142	1	IL3_SNGOE	P51445 saguinus o
23	38	55.1	249	1	NODH_RH1R	P52994 rhizobium
24	38	55.1	304	1	TH1R_YEAST	O08666 saccharomy
25	38	55.1	310	1	ASTL_CONTA	P42652 cectunrix c
26	38	55.1	538	1	121R_HUMAN	Q91865 homo sapien
27	38	55.1	844	1	SECA_STAFA	P47974 strapyloco
28	37	53.6	111	1	HSEB_DROME	P29978 drosophila
29	37	53.6	181	1	Y06E_BPT4	O13111 bacterioph
30	37	53.6	213	1	SRN2_YEAST	P59176 saccharomy
31	37	53.6	280	1	TH1M_ECOS7	P55388 escherichi
32	37	53.6	280	1	TH1M_ECOS1	P31142 escherichi
33	37	53.6	296	1	TH1R_BOVIN	P00586 bos taurus

45	36	52.2	185	1	YOL1_RHOER	G9Zg11	rhodococcus
44	37	53.6	746	1	EX11_MOUSE	P97464	mus musculus
43	37	53.6	746	1	EX11_HUMAN	Q16363	homo sapiens
42	37	53.6	746	1	EX11_CRIGR	C9Jk82	cricetulus
41	37	53.6	464	1	TRME_BORBU	P53364	borrelia bu
40	37	53.6	435	1	YINB_ECOLI	P78067	escherichia
39	37	53.6	337	1	AXF_YARLI	Q23839	yararowia ii
38	37	53.6	368	1	OGC_EWCCA	Q59483	erynia car
37	37	53.6	375	1	VNS5_MSTV	P16133	maire stripi
36	37	53.6	369	1	P111_HUMAN	P11188	homo sapien
35	37	53.6	341	1	YEJ8_ECOLI	P45490	campylobact
34	37	53.6	331	1	Y487_CAMEE	P55430	campylobact

## ALIGNMENTS

```

RESULT 1
ID      TH11 MYCBO      STANDARD;      FRT;      277 AA.
AC      PS9893;
DT      15-MAR-2004 (Rel. 43, Created)
DT      15-MAR-2004 (Rel. 43, Last sequence update)
DT      15-MAR-2004 (Rel. 43, Last annotation update)
DE      Putative thiosulfate sulfurtransferase 1 (EC 2.8.1.1) (Rhodanese-like
GN      protein 1).
OS      CYSAL OR MB0838C.
OC      Mycobacterium bovis.
OC      Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OX      Corynebacterineae; Mycobacteriaceae; Mycobacterium.
RN      NCBI_TaxID=1765;
RP      [1]
RZ      SEQUENCE FROM N.A.
RX      MEDLINE=22709107; PubMed=127865972;
RA      Garmier T., Elgmeyer K., Camus J.-C., Medina N., Mansoor H.,
RA      Pryor M., Duthoy S., Grondin S., Lacroix C., Monsempé C., Simon S.,
RA      Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA      Parkhill J., Barrett B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
RT      "The complete genome sequence of Mycobacterium bovis."
RL      Proc. Natl. Acad. Sci. U.S.A. 100:7877-7883(2003).
CC      -1- FUNCTION: May be a sulfotransferase involved in the formation of
CC      thiosulfate (by similarity).
CC      -1- CATALYTIC ACTIVITY: Thiosulfate + cyanide = sulfite + thiocyanate.
CC      -1- SIMILARITY: Contains 2 rhodanese domains.
CC      -----
CC      CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation
CC      at the European Bioinformatics Institute. There are no restrictions on
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for
CC      commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/sib-sib.ch).
CC      or send an email to license@sib-sib.ch).
CC      -----
DR      PROSITE; PS00380; RHODANESE_1; FALSE_NEG.
DR      PROSITE; PS00380; RHODANESE_2; 1.
DR      PROSITE; PS0206; RHODANESE_3; 2.
KW      Transferase; Repeat; Complete proteome.
FT      DOMAIN 18 125 RHODANESE 1.
FT      DOMAIN 154 274 RHODANESE 2.
FT      ACT SITE 233 233 BY SIMILARITY.
SQ      SEQUENCE 277 AA; 31014 MW; AC7B715D959565A9 CRC64;

Query Match 68.1%; Score 47; DB 1; Length 277;
Best Local Similarity 54.5%; Pred. No. 1,1;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

2 ISSYDGSWNDY 12
VKNIIDIIIIII:
255 VKNIIDGSWTEY 265

```

```

RESULT 2
THTR_MYCTU STANDARD; PRT; 277 AA.
ID THTR_MYCTU
AC 05036;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Putative thiosulfate sulfotransferase (EC 2.8.1.1) (Rhodanese-like protein)
GN CYSA OR CYSA3 OR ML2198.
OS Mycobacterium leprae.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1]
RP SEQUENCE FROM N.A.
RA Smith D.R., Robinson K.;
RL Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Cole S.T., Biglamer K., Parkhill J., James K.D., Thomson N.R., Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D., Mungall K., Basham D., Brown D., Chillingworth T., Connor R., Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N., Holtroyd S., Hornsby T., Jagsels K., Jancsik C., Maclean J., Moule S., Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M., Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R., Squires S., Stevens K., Taylor K., Whitehead S., Woodward J.R., Barrett B.G.;
RT "Massive gene decay in the leprosy bacillus."
RL Nature 409:1007-1011(2001).
CC -1- FUNCTION: May be a sulfotransferase involved in the formation of thiosulfate.
CC -1- CATALYTIC ACTIVITY: Thiosulfate + cyanide = sulfite + thiocyanate.
CC -1- SIMILARITY: Contains 2 rhodanese domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U15182; AAA62982.1; -
DR EMBL; AL583924; CAC31153.1; -
DR PIR; A87184; A87184.
DR Leproma; ML2198; -
DR InterPro; IPR001763; Rhodanese-like.
DR InterPro; IPR001507; Rhodanese.
DR Pfam; PF00581; Rhodanese; 2.
DR SMART; SM00450; RHOD; 2.
DR PROSITE; PS00380; RHODANES1; FALSE_NEG.
DR PROSITE; PS00683; RHODANES2; 1.
DR PROSITE; PS50206; RHODANES3; 2.
KW Transferrase; Complete proteome; Repeat.
FT DOMAIN 18 125 RHODANES1.
FT DOMAIN 154 274 RHODANES2.
FT ACT_SITE 233 233 BY SIMILARITY.
SQ SEQUENCE 277 AA; 31094 MW; 527E05CFE93969D CRC64;

Query Match 68.1%; Score 47; DB 1; Length 277;
Best local similarity 54.5%; Pred. No. 1.1;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

```

```

THTR_MYCTU STANDARD; PRT; 277 AA.
ID THTR_MYCTU
AC 005793;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Putative thiosulfate sulfotransferase (EC 2.8.1.1) (Rhodanese-like protein)
GN CYSA1 OR CYSA OR RVJ117 OR MTJ199 OR MTCY164.27 AND (CYSA2 OR RV0815C OR MT0837 OR MT0493.07C).
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=H37Rv;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D., Gordon S.V., Eigmler K., Gas S., Barry C.E. III, Tekala F., Badcock K., Basham D., Brown D., Chillingworth T., Connor R., Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holtroyd S., Hornsby T., Jagsels K., Krogh A., McLean J., Moule S., Murphy L., Oliver S., Osborne U., Quail M.A., Rajandream M.A., Rogers R., Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E., Taylor K., Whitehead S., Barrett B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence."
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RA STRAIN=CDC 1551 / Oshkosh;
RX MEDLINE=22206494; PubMed=12218036;
RA Fleischmann R.D., Alland D., Bisen J.A., Carpenter L., White O., Peterson J., DeBoy R., Dodson R., Gwin M., Hatt D., Hickey E., Kolony J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L., Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A., Bishai W., Jacobs W.R. Jr., Venter U.C., Fraser C.M.;
RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and laboratory strains."
RL J. Bacteriol. 184:5479-5490(2002).
CC -1- FUNCTION: May be a sulfotransferase involved in the formation of thiosulfate.
CC -1- CATALYTIC ACTIVITY: Thiosulfate + cyanide = sulfite + thiocyanate.
CC -1- SIMILARITY: Contains 2 rhodanese domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z95150; CAB08374.1; -
DR EMBL; AL022004; CAAL7621.1; -
DR EMBL; AE006973; AAK45079.1; -
DR EMBL; AE007135; AAK47535.1; -
DR PIR; G70809; G70809.
DR HSP; P00586; IRHS.
DR TIGR; MT0837; -
DR TIGR; MTJ199; -
DR Tuberculist; RV0815C; -
DR InterPro; IPR001763; Rhodanese-like.
DR InterPro; IPR001507; Rhodanese.
DR Pfam; PF00581; Rhodanese; 2.
DR SMART; SM00450; RHOD; 2.
DR PROSITE; PS00380; RHODANES1; FALSE_NEG.
DR PROSITE; PS00683; RHODANES2; 1.
DR PROSITE; PS50206; RHODANES3; 2.
KW Transferrase; Repeat; Complete proteome.
FT DOMAIN 18 125 RHODANES1.
FT DOMAIN 154 274 RHODANES2.

```

RESULT 3

```
QY      2 ISSYDGSWNDY 12  
          : :| | | | : |  
Db      255 VKNYDGSWTEY 265
```

OS Rattus norvegicus (Rat).

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-Strague-Dawley; TISSUE-Liver;  
 RX MEDLINE=91207296; PubMed=2018478;  
 RA Weiland K.L., Dooley T.F.;  
 RT "Molecular cloning, sequencing and characterization of cDNA to rat  
 RT liver rhodanese, a thiosulphate sulphurtransferase.";  
 RL Biochem. J. 275:227-231(1991).  
 RN [2]  
 RP MUTAGENESIS.  
 RC TISSUE-Liver;  
 RX MEDLINE=95332330; PubMed=7609189;  
 RA Nagahara N., Okazaki T., Nishino T.;  
 RT "Cytosolic mercaptopyruvate sulfurtransferase is evolutionarily  
 RT related to mitochondrial rhodanese. Striking similarity in active site  
 RT amino acid sequence and the increase in the mercaptopyruvate  
 RT sulfurtransferase activity of rhodanese by site-directed  
 RT mutagenesis.";  
 RL J. Biol. Chem. 270:16230-16235(1995).  
 CC -1- FUNCTION: INVOLVED IN THE FORMATION OF IRON-SULFUR COMPLEXES,  
 CC CYANIDE DETOXIFICATION OR MODIFICATION OF SULFUR-CONTAINING  
 CC ENZYMES. OTHER THIO COMPOUNDS, BESIDES CYANIDE, CAN ACT AS SULFUR  
 CC ION ACCEPTORS. ALSO HAS WEAK MERCAPTOPYRUVATE SULFURTRANSFERASE  
 CC (MST) ACTIVITY.  
 CC -1- CATALYTIC ACTIVITY: Thiosulfate + cyanide = sulfite + thiocyanate.  
 CC -1- SUBUNIT: Monomer.  
 CC -1- SUBCELLULAR LOCATION: Mitochondrial matrix.  
 CC -1- TISSUE SPECIFICITY: Expressed in numerous tissues.  
 CC -1- DOMAIN: THE STRUCTURE CONSISTS OF 2 DOMAINS OF VERY SIMILAR  
 CC CONFORMATION, SUGGESTING A COMMON EVOLUTIONARY ORIGIN. HOWEVER,  
 CC THE SEQUENCES OF THE 2 DOMAINS ARE VERY DIFFERENT.  
 CC -1- SIMILARITY: Contains 2 rhodanese domains.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL, X56228; CAA39677.1; -.  
 DR PIR, S15081; S15081.  
 DR HSSP, P00586; IRHS.  
 DR InterPro, IPR001763; Rhodanese-like.  
 DR InterPro, IPR001307; Rhodanese.  
 DR Pfam, PF00581; Rhodanese; 2.  
 DR SMART, SM00450; RHOD; 2.  
 DR PROSITE, PS00380; RHODANES\_1; 1.  
 DR PROSITE, PS00683; RHODANES\_2; 1.  
 DR PROSITE, PS50206; RHODANES\_3; 2.  
 KW Transferrase; Mitochondrion; Repeat.  
 FT NON TER 1 1  
 FT DOMAIN 23 141 RHODANES 1.  
 FT DOMAIN 142 157 HINGE.  
 FT DOMAIN 171 286 RHODANES 2.  
 FT ACT\_SITE 185 185 MAY PLAY A ROLE IN SUBSTRATE BINDING (BY  
 FT ACT\_SITE 185 185 SIMILARITY).  
 FT ACT\_SITE 246 246 BY SIMILARITY.  
 FT ACT\_SITE 247 247 SUBSTRATE (THIOSULFATE) BINDING.  
 FT ACT\_SITE 248 248 SUBSTRATE (THIOSULFATE) BINDING.  
 FT MUTAGEN 247 247 R->G: UNALTERED RHODANES ACTIVITY;  
 FT MUTAGEN 248 248 INCREASED MST ACTIVITY.  
 FT MUTAGEN 248 248 K->S: DECREASED RHODANES ACTIVITY;  
 FT FT UNALTERED MST ACTIVITY.  
 FT SEQUENCE 295 AA; 33176 MW; 24655B3569093451 CRC64;

QY 2 ISSYDGSNDY 12  
 Db 267 VAYYDGSMSSEW 277  
 RESULT 7  
 ID THF2\_MYCLE STANDARD; PRT; 296 AA.  
 AC P46700;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 26-FEB-2003 (Rel. 41, Last annotation update)  
 DE Putative thiosulfate sulfurtransferase ssea (EC 2.8.1.1).  
 GN SSEA OR M10728 OR B1508\_C1\_127.  
 OS Mycobacterium leprae.  
 OC Bacteria; Actinobacteria; Actinomycetales;  
 OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_TaxID=1769;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96084954; PubMed=7489918;  
 RA Doukhan L., Predich M., Nair G., Dussurget O., Mandic-Mulec I.,  
 RA Cole S.T., Smith D.R., Smith I.;  
 RT "Genomic organization of the mycobacterial sigma gene cluster.";  
 RL Gene 165:67-70(1995).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=TN;  
 RX MEDLINE=21128732; PubMed=11234002;  
 RA Cole S.T., Eiglmeyer K., Parkhill J., James K.D., Thomson N.R.,  
 RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,  
 RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,  
 RA Holooyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,  
 RA Murphy L., Oliver K., Quail M.A., Rajandream K.A., Rutherford K.M.,  
 RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,  
 RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,  
 RA Barrett B.G.;  
 RT "Massive gene decay in the leprosy bacillus.";  
 RL Nature 409:1007-1011(2001).  
 CC -1- CATALYTIC ACTIVITY: Thiosulfate + cyanide = sulfite + thiocyanate.  
 CC -1- SIMILARITY: Contains 2 rhodanese domains.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL, U00012; AAA85919.1; -.  
 DR EMBL, AL583919; CAC30237.1; -.  
 DR PIR, A87000; A87000.  
 DR HSSP, P52197; LBOC.  
 DR Leproma, M10728; -.  
 DR InterPro, IPR001763; Rhodanese-like.  
 DR InterPro, IPR001307; Rhodanese.  
 DR Pfam, PF00581; Rhodanese; 2.  
 DR SMART, SM00450; RHOD; 2.  
 DR PROSITE, PS00380; RHODANES\_1; 1.  
 DR PROSITE, PS00683; RHODANES\_2; 1.  
 DR PROSITE, PS50206; RHODANES\_3; 2.  
 KW Hypothetical protein; Transferrase; Complete proteome; Repeat.  
 FT DOMAIN 31 138 RHODANES 1.  
 FT DOMAIN 168 286 RHODANES 2.  
 FT ACT\_SITE 245 245 BY SIMILARITY.  
 FT SEQUENCE 296 AA; 33235 MW; D670A07D5010144 CRC64;

Query Match 59.4%; Score 41; DB 1; Length 295;  
 Best Local Similarity 45.5%; Pred. No. 12;  
 Matches 5; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Query Match 59.4%; Score 41; DB 1; Length 296;  
 Best Local Similarity 45.5%; Pred. No. 12;  
 Matches 5; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 2 ISSYDGSWMDY 12  
DB 267 VRYDGSWTEW 277

## RESULT 8

THTR\_CRIGR STANDARD; PRT; 296 AA.  
ID THTR\_CRIGR  
AC P46635;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Thiosulfate sulfurtransferase (EC 2.8.1.1) (Rhodanese).  
GN TST.  
OS Cricetus griseus (Chinese hamster).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;  
OC Cricetulus.  
OX NCBI\_TaxID=10029;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Ovary;  
RX MEDLINE=96106946; PubMed=8535164;  
RA Trevino R.J., Hunt J., Horowitz P.M., Chirgwin J.M.;  
RT "Chinese hamster rhodanese cDNA: activity of the expressed protein is  
not blocked by a C-terminal extension."  
RL Protein Expr. Purif. 6:693-699(1995).  
CC -1- FUNCTION: FORMATION OF IRON-SULFUR COMPLEXES AND CYANIDE  
DETOXIFICATION. BINDS MOLECULAR OXYGEN AND SULFUR.  
CC -1- CATALYTIC ACTIVITY: Thiosulfate + cyanide = sulfite + thiocyanate.  
CC -1- SUBUNIT: Monomer (By similarity).  
CC -1- SUBCELLULAR LOCATION: Mitochondrial matrix.  
CC -1- DOMAIN: THE STRUCTURE CONSISTS OF 2 DOMAINS OF VERY SIMILAR  
CONFORMATION SUGGESTING A COMMON EVOLUTIONARY ORIGIN. HOWEVER,  
THE SEQUENCES OF THE 2 DOMAINS ARE VERY DIFFERENT.  
CC -1- SIMILARITY: Contains 2 rhodanese domains.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: U23943; AAB84305.1; -  
DR HSSP: P00586; IRHS.  
DR InterPro: IR001763; Rhodanese-like.  
DR InterPro: IR001307; Rhodanese.  
DR Pfam: PF00581; Rhodanese; 2.  
DR SMART: SM00450; RHOD; 2.  
DR PROSITE: PS00380; RHODANES\_1; 1.  
DR PROSITE: PS00683; RHODANES\_2; 1.  
DR PROSITE: PS50206; RHODANES\_3; 2.  
KM Transferase; Mitochondrion; Repeat.  
FT INIT\_MET 0  
FT DOMAIN 24 142  
FT DOMAIN 143 158  
FT DOMAIN 172 287  
FT ACT\_SITE 186 186  
FT ACT\_SITE 247 247  
FT ACT\_SITE 248 248  
FT ACT\_SITE 249 249  
FT ACT\_SITE 296 AA; 33205 MW; 8811022B9BAD66 CRC64;  
SQ SEQUENCE

Query March 59.4%; Score 41; DB 1; Length 296;  
Beet local Similarity 45.5%; Pred. No. 12;  
Matches 5; Conservative 5; Mismatches 1; Indels 0; Gaps 0;  
QY 2 ISSYDGSWMDY 12

DB 268 VAVYDGSWSEW 278

## RESULT 9

THTR\_HUMAN STANDARD; PRT; 296 AA.  
ID THTR\_HUMAN  
AC Q16762;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Thiosulfate sulfurtransferase (EC 2.8.1.1) (Rhodanese).  
GN TST.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Fetal liver;  
RX MEDLINE=97223398; PubMed=9070219;  
RA Alfa N., Ishii K., Akamatsu Y., Ogasawara Y., Tanabe S.;  
RT "Cloning and expression of human liver rhodanese cDNA."  
RL Biochem. Biophys. Res. Commun. 231:56-60(1997).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20057165; PubMed=10591208;  
RA Dunham I., Hunt A.R., Collins J.E., Brunskewich R., Beare D.M.,  
Clamp M., Smink L.J., Ainscough R., Almeida J.P., Babbage A.K.,  
Baggaley C., Bailey U., Barlow K.F., Bates K.N., Beasley O.P.,  
Bird C.P., Blakey S.E., Bridgman A.M., Buck D., Burgess J.,  
Burhill W.D., Burton J., Carder C., Carter N.P., Chen Y., Clark G.,  
Clegg S.M., Cobley V.E., Cole C.G., Collier R.E., Connor R.,  
Conroy D., Corby N.R., Coville G.J., Cox A.V., Davis J., Dawson E.,  
Dhami P.D., Dockree C., Dodsorth S.J., Dublin R.M., Ellington A.G.,  
Evans K.L., Fey J.M., Fleming K., French L., Garner A.A.,  
Gilbert J.G.R., Goward M.E., Grahm D.V., Griffiths M.N.D., Hall C.,  
Hall R.E., Hall-Tamlyn G., Heathcote R.W., Ho S., Holmes S.,  
Hunt S.E., Jones M.C., Kershaw J., Kimberley A.W., King A.,  
Laird G.K., Langford C.F., Leverisha M.A., Lloyd C., Lloyd D.M.,  
Marx I.D., Mashreghi-Mohammadi M., Matthews L.H., Mccann O.T.,  
McCarthy J., McLaren S., McMuray A.A., Milne S.A., Mortimore B.J.C.T.,  
Odell C.N., Pavitt R., Pearce A.V., Pearson D., Phillimore B.J.C.T.,  
Phillips S.H., Plumb R.W., Ramsey H., Ramsey Y., Rogers L., Ross M.T.,  
Scott C.E., Selha H.K., Skuce C.D., Smalley S., Smith M.L.,  
Soderlund C., Spragon L., Steward C.A., Sulston J.E., Swann R.M.,  
Vaudin M., Wall M., Wallis J.M., Whiteley M.N., Willey D.L.,  
Williams L., Williams S.A., Williamson H., Wilmer T.E., Wilting L.,  
Wright C.L., Hubbard T., Bentley D.R., Beck S., Rogers J., Shimizu N.,  
Minoshima S., Kawasaki K., Sasaki T., Asakawa S., Kudoh J.,  
Shintani A., Shibuya K., Yoshizaki Y., Aoki N., Mitsuyama S.,  
Roe B.A., Chen F., Chu J., Crabtree J., Deschamps S., Do A., Do T.,  
Dorman A., Fang F., Fu Y., Hu P., Hua A., Kenton S., Lai H., Lao H.I.,  
Lewis J., Lewis S., Lin S.-P., Loh P., Malaj E., Nguyen T., Pan H.,  
Phan S., Qi S., Qian Y., Ray L., Ren Q., Shauli S., Sloan D., Song L.,  
Wang Q., Wang Y., Wang Z., White J., Willingham D., Wu H., Yao Z.,  
Zhan M., Zhang G., Chisase S., Murray J., Miller N., Minx P.,  
Fulton R., Johnson D., Semis G., Bentley D., Bradshaw H., Bourne S.,  
Cordes M., Du Z., Fulton L., Goela D., Graves T., Hawkins J., Kohling T.,  
Hinds K., Kemp K., Latreille P., Layman D., Ozersky P., Rohlfing T.,  
Rasche P., Walker C., Wamsley A., Wohlmann P., Pepin K., Nelson J.,  
Korf I., Bedell J.A., Hillier L.W., Mardis E., Waterston R.,  
Wilson R., Emanuel B.S., Shaikh T., Kurahashi H., Saitta S.,  
Budarf M.L., McDermid H.E., Johnson A., Wong A.C.C., Morrow B.E.,  
Edelmann L., Kim U.J., Shiruya H., Simon M.I., Dunanski J.P.,  
Reyraud M., Kedra D., Seroussi E., Franson I., Tepia I., Bruder C.E.,  
O'Brien K.P., Wilkinson P., Bodenreich A., Hartman K., Hu X.,  
Khan A.S., Lane L., Tilihan Y., Wright H.;  
RT "The DNA sequence of human chromosome 22."  
RL Nature 402:489-495 (1999).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC TISSUE=uterus;

RA MEDLINE=22388257; PubMed=12477932;  
 RA Strassberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusier K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stabileon M., Soares M.B., Bonaldi W.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Ushin T.B., Toshiyuki S., Carrinot P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
 RA Rask S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Villalón D.K., Muzny K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W.,  
 RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson W.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallus D.E.,  
 RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.,  
 RA "Generation and initial analysis of more than 15,000 full-length  
 RT human and mouse cDNA sequences".  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 [4]  
 RN SEQUENCE OF 1-7.  
 RP TISSUE=Platelet;  
 RX MEDLINE=22608298; PubMed=12655801;  
 RA Gevaert K., Goethals M., Martens L., Van Damme J., Staes A.,  
 RA Thomas G.R., Vandekerckhove J.,  
 RT "Exploring proteomes and analyzing protein processing by mass  
 RT spectrometric identification of sorted N-terminal peptides.";  
 RL Nat. Biotechnol. 21:566-569(2003).  
 CC -1- FUNCTION: FORMATION OF IRON-SULFUR COMPLEXES, CYANIDE  
 CC DETOXIFICATION OR MODIFICATION OF SULFUR-CONTAINING ENZYMES.  
 CC OTHER THIOL COMPOUNDS, BESIDES CYANIDE. CAN ACT AS SULFUR ION  
 CC ACCEPTORS. ALSO HAS WEAK MERCAPTOPYRUVATE SULFURTRANSFERASE (MST)  
 CC ACTIVITY (BY SIMILARITY).  
 CC -1- CATALYTIC ACTIVITY: Thiosulfate + cyanide = sulfite + thiocyanate.  
 CC -1- SUBUNIT: Monomer.  
 CC -1- SUBCELLULAR LOCATION: Mitochondrial matrix.  
 CC -1- DOMAIN: THE STRUCTURE CONSISTS OF 2 DOMAINS OF VERY SIMILAR  
 CC CONFORMATION, SUGGESTING A COMMON EVOLUTIONARY ORIGIN. HOWEVER,  
 CC THE SEQUENCES OF THE 2 DOMAINS ARE VERY DIFFERENT.  
 CC -1- SIMILARITY: Contains 2 rhodanese domains.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; D87292; BA13327.1; -  
 DR EMBL; Z73420; CA97762.1; -  
 DR EMBL; BC010148; AAH10148.1; -  
 DR PIR; J05286; J05286.  
 DR HSSP; P00586; IRRS.  
 DR GeneW; HGNC:12388; TST.  
 DR MIM; 180370; -  
 DR GO; GO:0004792; F:thiosulfate sulfurtransferase activity; TAS.  
 DR GO; GO:0009440; P:cyanate catabolism; TAS.  
 DR InterPro; IPR001763; Rhodanese-like.  
 DR InterPro; IPR001307; Rhodanese.  
 DR Pfam; PF00581; Rhodanese; 2.  
 DR SMART; SM00450; RHOD; 2.  
 DR PROSITE; PS00380; RHODANES\_1; 1.  
 DR PROSITE; PS00683; RHODANES\_2; 1.  
 DR PROSITE; PS50206; RHODANES\_3; 2.  
 DR TRANSFERASE; Mitochondrion; Repeat.  
 DR INT MET 0  
 FT DOMAIN 24 142 RHODANES 1.  
 FT DOMAIN 143 158 HINGE.  
 FT DOMAIN 172 287 RHODANES 2.  
 FT ACT\_SITE 186 186 MAY PLAY A ROLE IN SUBSTRATE BINDING (BY

FT ACT\_SITE 247 247 SIMILARITY).  
 FT ACT\_SITE 248 248 BY SIMILARITY.  
 FT ACT\_SITE 249 249 SUBSTRATE (THIOSULFATE) BINDING  
 FT ACT\_SITE 249 249 (BY SIMILARITY).  
 FT ACT\_SITE 249 249 SUBSTRATE (THIOSULFATE) BINDING  
 FT ACT\_SITE 249 249 (BY SIMILARITY).  
 SQ SEQUENCE 296 AA; 33297 MW; 872C52008AEBDC5B CRC64;  
 Query Match 59.4%; Score 41; DB 1; Length 296;  
 Best Local Similarity 45.5%; Pred. No. 12;  
 Matches 5; Conservative 5; Mismatches 1; Indels 0; Gaps 0;  
 QY 2 ISSYDQSNQY 12  
 DB 268 VAVYDQSNQY 278  
 RESULT 10  
 THR\_MOUSE  
 ID THR\_MOUSE STANDARD; PRT; 296 AA.  
 AC P52196;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Thiosulfate sulfurtransferase (EC 2.8.1.1) (Rhodanese).  
 GN TST.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_Taxid=10090;  
 RX [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BALB/c; TISSUE=Liver;  
 RX MEDLINE=96074596; PubMed=7488186;  
 RA Dooley T.P., Nair S.K., Garcia R.E., Courtney B.C.,  
 RT "Mouse rhodanese gene (Tst): cDNA cloning, sequencing, and  
 RT recombinant protein expression.";  
 RL Biochem. Biophys. Res. Commun. 216:1101-1109(1995).  
 CC -1- FUNCTION: FORMATION OF IRON-SULFUR COMPLEXES AND CYANIDE  
 CC DETOXIFICATION.  
 CC -1- CATALYTIC ACTIVITY: Thiosulfate + cyanide = sulfite + thiocyanate.  
 CC -1- SUBUNIT: Monomer.  
 CC -1- SUBCELLULAR LOCATION: Mitochondrial matrix.  
 CC -1- TISSUE SPECIFICITY: Expressed in numerous tissues.  
 CC -1- DOMAIN: THE STRUCTURE CONSISTS OF 2 DOMAINS OF VERY SIMILAR  
 CC CONFORMATION, SUGGESTING A COMMON EVOLUTIONARY ORIGIN. HOWEVER,  
 CC THE SEQUENCES OF THE 2 DOMAINS ARE VERY DIFFERENT.  
 CC -1- SIMILARITY: Contains 2 rhodanese domains.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; U35741; AAC52342.1; -  
 DR PIR; JC4398; JC4398.  
 DR HSSP; P00586; IRRS.  
 DR SWISS-2DPAGE; P52196; MOUSE.  
 DR MGD; MGI:98852; Tst.  
 DR InterPro; IPR001763; Rhodanese-like.  
 DR InterPro; IPR001307; Rhodanese.  
 DR Pfam; PF00581; Rhodanese; 2.  
 DR SMART; SM00450; RHOD; 2.  
 DR PROSITE; PS00380; RHODANES\_1; 1.  
 DR PROSITE; PS00683; RHODANES\_2; 1.  
 DR PROSITE; PS50206; RHODANES\_3; 2.  
 DR TRANSFERASE; Mitochondrion; Repeat.  
 DR INT MET 0  
 FT DOMAIN 24 142 RHODANES 1.  
 FT DOMAIN 143 158 HINGE.

```

DOMAIN      172      287      RHODANSE 2.
FT ACT_SITE 186      186      MAY PLAY A ROLE IN SUBSTRATE BINDING (BY
FT ACT_SITE 247      247      SIMILARITY).
FT ACT_SITE 248      248      BY SIMILARITY.
FT ACT_SITE 249      249      SUBSTRATE (THIOSULFATE) BINDING
FT ACT_SITE 249      249      (BY SIMILARITY).
FT ACT_SITE 249      249      SUBSTRATE (THIOSULFATE) BINDING
FT ACT_SITE 249      249      (BY SIMILARITY).
SEQUENCE    296 AA; 33334 MM; 82089D80F9A8E55A CRC64;

Query Match      59.4%; Score 41; DB 1; Length 296;
Best Local Similarity 45.5%; Pred. No. 12;
Matches 5; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

OY      2      ISSYDGSMDNY 12
Db      268      VAVYDGSMSSEW 278

RESULT 11
THT2_MYCTU      STANDARD;      PRT,      297 AA.
AC P96888;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Putative thiosulfate sulfotransferase SSEA (BC 2.8.1.1).
GN SSEA OR RV3283 OR MT3382 OR MCTY1.23.
OS Mycobacterium tuberculosis.
CC Bacteria; Actinobacteriae; Actinobacteridae; Actinomycetales;
CC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
CX NCBI_TaxID=1773;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Broesch R., Parkhill J., Garner T., Churcher C., Harris D.,
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Dahlen K., Fellwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagers K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrall B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RA complete genome sequence."
RA Nature 393:537-544 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ODC 1551 / Oshkosh;
RX MEDLINE=22206494; PubMed=12218036;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M., Hart D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA Delcher A., Utechack T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Biswal M., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RA "Whole-genome comparison of Mycobacterium tuberculosis clinical and
RA laboratory strains."
RL J. Bacteriol. 184:5479-5490(2002).
CC -1- CATALYTIC ACTIVITY: Thiosulfate + cyanide = sulfite + thiocyanate.
CC -1- SIMILARITY: Contains 2 rhodanase domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z92771; CAB07066.1; -.
DR EMBL; AE007147; AAK47725.1; -.
DR F1R; D70980; D70980.

```

```

DR HSSP; P5219; 1E0C.
DR TIGR; MT3382; -.
DR TubercuList; RV3283; -.
DR InterPro; IPR001763; Rhodanese-like.
DR InterPro; IPR001307; Rhodanese.
DR Pfam; PF00581; Rhodanese; 2.
DR SMART; S000450; RHOD; 2.
DR PROSITE; PS00380; RHODANESE_1; 1.
DR PROSITE; PS00683; RHODANESE_2; 1.
DR PROSITE; PS50206; RHODANESE_3; 2.
DR Hypothetical protein; Transferrase; Complete proteome; Repeat.
FT DOMAIN 31 138 RHODANESE 1.
FT ACT_SITE 168 286 RHODANESE 2.
FT ACT_SITE 245 245 BY SIMILARITY.
SQ SEQUENCE 297 AA; 33320 MW; 5930AB5F9C63A33 CRC64;

Query Match 59.4%; Score 41; DB 1; Length 297;
Best Local Similarity 45.5%; Pred. No. 12;
Matches 5; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 2 ISSYDGSMDY 12
: ||||| :
Db 267 VRNYDGSMTWM 277

RESULT 12
Y4WH_RHISN STANDARD; PRT; 297 AA.
1D Y4WH_RHISN
AC P55567;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Hypothetical 33.9 kDa protein Y4WH.
GN Y4WH.
OS Rhizobium sp. (strain NGR234).
OG Plasmid sym NGR234a.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Rhizobium/Agrobacterium group; Rhizobium.
CX NCB1_TaxID=394;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97305956; PubMed=9163424;
RA Freiberg C.A., Fellay R., Bairoch A., Broughton W.J., Rosenthal A.,
RA Perret X.;
RT "Molecular basis of symbiosis between Rhizobium and legumes.";
RL Nature 387:394-401(1997).
CC -1- SIMILARITY: None obvious.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB000085; AAB91771.1; -
DR InterPro; IPR006992; Amidohydro_2.
DR Pfam; PF004909; Amidohydro_2; 1.
KW Hypothetical protein; Plasmid.
SQ SEQUENCE 297 AA; 33885 MW; 4DE2324557B63131 CRC64;

Query Match 59.4%; Score 41; DB 1; Length 297;
Best Local Similarity 58.3%; Pred. No. 12;
Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 ISSYDGSMDY 12
: ||||| :
Db 260 LFSSYDAIMNAP 271

RESULT 13
IL3B_MOUSE

```

```

ID IL3B_MOUSE STANDARD; PRT; 878 AA.
AC P26954;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Interleukin-3 receptor class II beta chain precursor (Colony
DE stimulating factor 2 receptor, beta 2 chain).
GN CSF2RB2 OR AIZCA OR IL3RB2 OR IL3R.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90117145; PubMed=2404337;
RA Itoh N., Yonehara S., Schreurs J., Gorman D.M., Maruyama K., Ishii A.,
RA Yahara I., Arai K., Miyajima A.;
RT Cloning of an interleukin-3 receptor gene: a member of a distinct
RT receptor gene family.
RL Science 247:324-327(1990).
CC -1- FUNCTION: In mouse, there are two classes of high-affinity IL3
CC receptors. One contains this IL3-specific beta chain and the other
CC contains the beta chain also shared by high-affinity IL5 and GM-
CC CSF receptors.
CC -1- SUBUNIT: Heterodimer of an alpha and a beta chain.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- SIMILARITY: Belongs to the type I cytokine family of receptors.
CC Subfamily 4.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M29855; AAA39295.1; .
DR PIR; A40091; A40091.
DR MGD; MGI:1339760; Csf2rb2.
DR InterPro; IPR002896; CR1A.
DR InterPro; IPR000282; Cytok_receptor_2.
DR InterPro; IPR008957; FN_III-like.
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR003531; Hemtopopn_S_F1.
DR Pfam; PF00041; fn3; 2.
DR SMART; PS00060; FN3; 2.
DR PROSITE; PS01355; HEMATOPO_REC_S_F1; 1.
DR Receptor; Transmembrane; Glycoprotein; Signal.
KM SIGNAL
FT 1 22
FT CHAIN
FT 23 878
FT DOMAIN 23 440
FT TRANSMEM 441 462
FT DOMAIN 463 878
FT DISULFID 39 49
FT DISULFID 78 95
FT DISULFID 254 264
FT DISULFID 310 310
FT CARBOHYD 62 62
FT CARBOHYD 350 350
FT SEQUENCE 878 AA; 97195 MW; 8BEC9092ADC24D56 CRC64;
SQ
Query Match 59.4%; Score 41; DB 1; Length 878;
Best Local Similarity 54.5%; Pred. No. 37;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
Oy 2 ISSYDGSWMDY 12
Db 420 ISYDGIWSEW 430
RESULT 14

```

```

CH1D_VIBFU STANDARD; PRT; 1046 AA.
AC P6156;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Chitodextrinase precursor (EC 3.2.1.14).
GN ENDO 1.
OS Vibrio furnissii.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrrio.
OX NCBI_TaxID=29494;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 31-51.
RX MEDLINE=97125982; PubMed=8969204;
RA Keyhani N.O., Roseman S.;
RT "The chitin catabolic cascade in the marine bacterium Vibrio
RT furnissii. Molecular cloning, isolation, and characterization of a
RT periplasmic chitodextrinase."
RL J. Biol. Chem. 271:33414-33424(1996).
CC -1- FUNCTION: Hydrolyzes chitin oligosaccharides; (GlcNAc) 4 to
CC (GlcNAc) 2 and (GlcNAc) 5, 6 to (GlcNAc) 2 and (GlcNAc) 3. Inactive
CC towards chitin, glucosamine oligosaccharides, glycoproteins and
CC glycopeptides containing (GlcNAc) 2. Has optimum activity at pH
CC 6.5-7.0 and at a temperature of 35-37 degrees Celsius.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of the 1,4-beta-linkages of N-
CC acetyl-D-glucosamine polymers of chitin.
CC -1- ENZYME REGULATION: Inhibited by (GlcNAc) 4, (GlcNAc) 5, (GlcNAc) 6,
CC and PNP-(GlcNAc) 3.
CC -1- PATHWAY: Chitin catabolism.
CC -1- SUBCELLULAR LOCATION: Periplasmic (probable).
CC -1- INDUCTION: BY (GlcNAc) 2.
CC -1- SIMILARITY: Belongs to chitinase class II (family 18 of glycosyl
CC hydrolases).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U41418; AAC44673.1; .
DR PIR; T30199; T30199.
DR InterPro; IPR003610; CEM_5_12.
DR InterPro; IPR001223; Glyco_hydro_18.
DR InterPro; IPR001579; Glyco_hydro_18AS.
DR InterPro; IPR000437; Prok_1lipoprot_S.
DR Pfam; PF02839; CEM_5_12; 2.
DR Pfam; PF00704; Glyco_hydro_18; 1.
DR PRODOM; PD000471; Glyco_hydro_18; 1.
DR SMART; SM00495; ChitD3; 2.
DR SMART; SM00636; Glyco_18; 1.
DR PROSITE; PS01095; CHITINASE_18; FALSE NEG.
KM Chitin degradation; Hydrolase; Glycosidase; Periplasmic; Signal.
FT SIGNAL
FT 1 30
FT CHAIN
FT 31 1046
FT SEQUENCE 1046 AA; 112380 MW; 40680F164D55A1F CRC64;
SQ
Query Match 59.4%; Score 41; DB 1; Length 1046;
Best Local Similarity 61.5%; Pred. No. 45;
Matches 8; Conservative 2; Mismatches 1; Indels 2; Gaps 1;
Oy 2 ISSYD-GSWMDY 12
Db 557 IMSYDHOAMDH 569
RESULT 15
YAN4_SCHPO STANDARD; PRT; 601 AA.
AC Q10070;

```



DT 01-FEB-1996 (Rel. 33, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Hypothetical protein C3H1.04c in chromosome I.  
GN SPAC3H1.04c.  
OS Schizosaccharomyces pombe (fission yeast).  
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
OC Schizosaccharomycetes.  
OX NCBI TaxID=4896;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=972;  
RX MEDLINE=2169401; PubMed=11859360;  
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,  
RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,  
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,  
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,  
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,  
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagsels K.,  
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,  
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,  
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,  
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
RA Skelton J., Simmonds M., Squares R., Stevens K., Stevens K.,  
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
RA Woodward J., Wolckaert G., Aert R., Robben J., Grymoprez B.,  
RA Welljens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,  
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,  
RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,  
RA Eger P., Zimmermann W., Wedler H., Wandutt R., Fumelle B.,  
RA Goffeau A., Gadiou E., Dreano S., Gloux S., Lelaure V., Mottier S.,  
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,  
RA Lucas M., Rocher M., Gallardin C., Tallada V.A., Garzon A., Thode G.,  
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
RA Dominguez A., Revela U.L., Moreno S., Armstrong C., Forsburg S.L.,  
RA Cerretti L., Lowe T., McCombie W.R., Paulsen I., Potashkin U.,  
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.,  
RT "The genome sequence of Schizosaccharomyces pombe.";  
RL Nature 415:871-880(2002).  
CC -!- SIMILARITY: TO YEAST YHR194W.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC  
DR EMBL: Z68144; CA92257.1; -.  
DR PIR: T38736; T38736.  
DR GeneDB spombe: SPAC3H1.04c; -.  
KM Hypothetical protein.  
SQ SEQUENCE 601 AA; 69428 MM; 8E2857C8A35DB72E CRC64;  
QY 2 ISSYDGSWMDY 12  
DB 528 LSDPDGSMWTF 538  
Query Match 58.0%; Score 40; DB 1; Length 601;  
Best Local Similarity 54.5%; Pred. No. 36;  
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Search completed: October 5, 2004, 08:01:49  
Job time : 2.07407 secs

This Page Blank (uspo1c)

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2004, 07:34:11 ; Search time 6.35185 seconds  
(without alignments)  
596.081 Million cell updates/sec

Title: US-09-805-290A-15  
Perfect score: 69  
Sequence: 1 LISSYDGSWMDY 12

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

SPTREMBL\_25:\*

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phage:\*
- 10: sp\_plant:\*
- 11: sp\_proteob:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_tvirus:\*
- 16: sp\_bacterioph:\*
- 17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	49	71.0	293	16	067668
2	49	71.0	708	5	087688
3	47	68.1	279	16	082661
4	47	68.1	281	16	082661
5	46	66.7	248	16	082661
6	45	65.2	332	11	082661
7	45	65.2	342	11	082661
8	45	65.2	361	11	082661
9	45	65.2	466	10	081876
10	45	65.2	466	10	081876
11	45	65.2	773	10	081876
12	45	65.2	779	10	081876
13	45	65.2	794	10	081876
14	45	65.2	804	10	081876
15	45	65.2	844	10	081876
16	45	65.2	889	10	081876

17	45	65.2	903	10	082661
18	45	65.2	929	10	082661
19	45	65.2	938	10	082661
20	45	65.2	955	10	082661
21	45	65.2	1110	10	082661
22	45	65.2	1153	10	082661
23	45	65.2	1281	10	082661
24	45	65.2	1286	10	082661
25	45	65.2	1353	10	082661
26	45	65.2	1405	10	082661
27	45	65.2	1415	5	061442
28	45	65.2	1421	10	07XHE0
29	45	65.2	1468	10	07XHE0
30	45	65.2	1477	10	07XHE0
31	45	65.2	1489	10	07XHE0
32	45	65.2	1512	10	07XHE0
33	45	65.2	1536	10	081N97
34	45	65.2	1553	10	07X924
35	45	65.2	1557	10	0855Y6
36	45	65.2	1590	10	07XHE2
37	45	65.2	1591	10	091DW9
38	45	65.2	1592	10	081Q12
39	45	65.2	1597	10	0855M7
40	45	65.2	1597	10	08H904
41	45	65.2	1597	10	07XN59
42	45	65.2	1597	10	07XHE1
43	45	65.2	1602	10	07XHE3
44	45	65.2	1604	10	08W062
45	45	65.2	1605	10	081R53

## ALIGNMENTS

RESULT 1

ID 067668 PRELIMINARY; PRT; 293 AA.

AC 067668;

DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)

DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)

DT 01-UN-2003 (TREMBLrel. 24, Last annotation update)

DE Thiosulfate sulfurtransferase.

GN RHD2 OR AQ.1199.

OS Aquifex aeolicus.

OC Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.

OX NCBI\_TaxID=63363;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=VF5;

RX MEDLINE=98196666; PubMed=9537320;

RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,

RA Graham D.B., Overbeek R., Sneed M.A., Keller M., Aufay M., Huber R.,

RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;

RT "The complete genome of the hyperthermophilic bacterium Aquifex

RT aeolicus";

RT Nature 392:353-358(1998).

RL EMBL; AE000757; AAC07633.1; -.

DR PIR; B70455; B70455.

DR GO; GO:0004792; F:thiosulfate sulfurtransferase activity; IEA.

DR GO; GO:0016740; F:transferase activity; IEA.

DR GO; GO:0008272; P:sulfate transport; IEA.

DR InterPro; IPR001307; Rhodanese.

DR InterPro; IPR001763; Rhodanese-like.

DR Pfam; PF00581; Rhodanese; 2.

DR SMART; SM00450; RHOD; 2.

DR PROSITE; PS00380; RHODANES\_1; 1.

DR PROSITE; PS00683; RHODANES\_2; 1.

DR PROSITE; PS02066; RHODANES\_3; 2.

DR TRANSFERASE; Complete proteome.

DR TRANSFERASE; 293 AA; 34468 MW; 44FC38A35244B383 CRC64;

SEQUENCE

Query Match 71.0%; Score 49; DB 16; Length 293;

Best Local Similarity 87.5%; Pred. No. 5.8;

Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 5 YDGSMDY 12  
DB 277 YDGSMDY 284

## RESULT 2

08T6B8 PRELIMINARY; PRT; 708 AA.  
AC 08T6B8;  
DT 01-JUN-2002 (TREMBlrel. 21, Created)  
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)  
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
DE Non-transporter ABC protein AbcF.  
GN ABCF.  
OS Dicyostelium discoideum (Slime mold).  
OC Eukaryota; Mycetozoa; Dicyostellida; Dicyostelium.  
OX NCBI\_Taxid=44689;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Ax4;  
RA Anjaud C., Loomis W.F.;  
RT "Evolution of the ABC transporters of Dicyostelium";  
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF479253; AAL87691.1; -  
DR GO; GO:0016020; C:membrane; IEA.  
DR GO; GO:0005524; F:ATP binding; IEA.  
DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. .; IEA.  
DR GO; GO:0000166; P:nucleotide binding; IEA.  
DR GO; GO:0006810; P:transport; IEA.  
DR InterPro; IPR003593; AAA\_ATPase.  
DR InterPro; IPR003439; ABC\_transporter.  
DR Pfam; PF00005; ABC\_tran; 2.  
DR ProDom; PD000006; ABC transporter; 2.  
DR SMART; SM00382; AAA; 2.  
DR PROSITE; PS00211; ABC\_TRANSPORTER\_1; 2.  
DR PROSITE; PS00893; ABC\_TRANSPORTER\_2; 2.  
DR KX AT-binding.  
SQ SEQUENCE 708 AA; 79895 MW; 9EF153773A354F05 CRC64;

Query Match 71.0%; Score 49; DB 5; Length 708;  
Best Local Similarity 54.5%; Pred. No. 15;  
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 2 ISSYDGSMDY 12  
DB 691 VKNFDGMDY 701

## RESULT 3

082G61 PRELIMINARY; PRT; 279 AA.  
AC 082G61;  
DT 01-JUN-2003 (TREMBlrel. 24, Created)  
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)  
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
DE Putative thiosulfate sulfurtransferase.  
GN CysA2 OR SAV4037.  
OS Streptomyces avermitilis.  
OC Bacteria; Actinobacteriae; Actinomycetales;  
OC Streptomycinae; Streptomycetaceae; Streptomyces.  
OX NCBI\_Taxid=33903;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;  
RA MEDLINE=21477403; PubMed=11572948;  
RA Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,  
RA Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osone T.,  
RA Kikuchi H., Shiba T., Sakaki Y., Hattori M.;  
RT "Genome sequence of an industrial microorganism Streptomyces  
metabolites";

RL Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).

RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;  
RX MEDLINE=22608306; PubMed=12692562;  
RA Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,  
RA Sakaki Y., Hattori M., Omura S.;  
RT "Complete genome sequence and comparative analysis of the industrial  
microorganism Streptomyces avermitilis";  
RL Nat. Biotechnol. 21:526-531(2003).  
DR EMBL; AP005037; BAC71749.1; -  
DR GO; GO:0004792; F:thiosulfate sulfurtransferase activity; IEA.  
DR GO; GO:0016740; F:transferase activity; IEA.  
DR GO; GO:0008272; P:sulfate transport; IEA.  
DR InterPro; IPR001307; Rhodanese-like.  
DR InterPro; IPR001763; Rhodanese-like.  
DR Pfam; PF00581; Rhodanese; 2.  
DR SMART; SM00450; RHOD; 2.  
DR PROSITE; PS00683; RHODANES\_2; 1.  
DR PROSITE; PS0206; RHODANES\_3; 2.  
DR KX Transferase; Complete proteome.  
SQ SEQUENCE 279 AA; 31592 MW; 3AAC72C789A0943 CRC64;

Query Match 68.1%; Score 47; DB 16; Length 279;  
Best Local Similarity 54.5%; Pred. No. 12;  
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 2 ISSYDGSMDY 12  
DB 255 VKNYDGSMDY 265

## RESULT 4

08CTJ5 PRELIMINARY; PRT; 281 AA.  
AC 08CTJ5;  
DT 01-MAR-2003 (TREMBlrel. 23, Created)  
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)  
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
DE Putative thiosulfate sulfurtransferase.  
GN SC04164 OR SCD66.01 OR SCD84.31.  
OS Streptomyces coelicolor.  
OC Bacteria; Actinobacteriae; Actinomycetales;  
OC Streptomycinae; Streptomycetaceae; Streptomyces.  
OX NCBI\_Taxid=1902;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A3(2) / M145;  
RX MEDLINE=21996410; PubMed=12000953;  
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,  
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kleser H.,  
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,  
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,  
RA Huang C.-H., Kleser T., Larke L., Murphy L., Oliver K., O'Neill S.,  
RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,  
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,  
RA Warren T., Wietzorek A., Woodward J., Barrell B.G., Parkhill J.,  
RA Hopwood D.A.;  
RT "Complete genome sequence of the model actinomycete Streptomyces  
coelicolor A3(2)";  
RL Nature 417:141-147(2002).  
DR EMBL; AL939119; CAD55474.1; -  
DR GO; GO:0004792; F:thiosulfate sulfurtransferase activity; IEA.  
DR GO; GO:0016740; F:transferase activity; IEA.  
DR GO; GO:0008272; P:sulfate transport; IEA.  
DR InterPro; IPR001307; Rhodanese-like.  
DR InterPro; IPR001763; Rhodanese-like.  
DR Pfam; PF00581; Rhodanese; 2.  
DR SMART; SM00450; RHOD; 2.  
DR PROSITE; PS00683; RHODANES\_2; 1.  
DR PROSITE; PS0206; RHODANES\_3; 2.  
DR KX Transferase; Complete proteome.  
SQ SEQUENCE 281 AA; 31771 MW; B72E4DA7735C790B CRC64;

```

Query Match      68.1%; Score 47; DB 16; Length 281;
Best Local Similarity 54.5%; Pred. No. 12;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY      2 ISSYDGSWMDY 12
      : ||||| :
DB      257 VKMYDGSWTEY 267

RESULT 5
Q8FOP3  PRELIMINARY; PRT; 248 AA.
AC      08FOP3;
DT      01-MAR-2003 (TREMBlrel. 23, Created)
DT      01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DE      Conserved hypothetical protein.
GN      CE1076.
OS      Corynebacterium efficiens.
OC      Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC      Corynebacterineae; Corynebacteriaceae; Corynebacterium.
OX      NCBI_TaxID=152794;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=YS-314 / AJ 12310 / DSM 44549 / JCM 11189;
RA      Kawababayasi Y., Yamazaki J., Hino Y., Kikuchi H., Nakamura Y.,
RA      Ikeo K., Suzuki M., Mashima J., Itoh T., Yamagishi A., Nishio Y.,
RA      Usuda Y., Sugimoto S.;
RT      The entire genome sequence of Corynebacterium efficiens YS-314;
RL      Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
KW      Hypothetical protein; Complete proteome.
SQ      SEQUENCE 248 AA; 26598 MW; B2AD266FBA7FD589 CRC64;

Query Match      66.7%; Score 46; DB 16; Length 248;
Best Local Similarity 54.5%; Pred. No. 15;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY      2 ISSYDGSWMDY 12
      : ||||| :
DB      177 VEGYDGSWMDY 187

RESULT 6
Q8JZNI  PRELIMINARY; PRT; 332 AA.
AC      08JZNI;
DT      01-OCT-2002 (TREMBlrel. 22, Created)
DT      01-OCT-2002 (TREMBlrel. 22, Last sequence update)
DT      01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE      Macrophage galactose-type C-type lectin 2.
GN      MGL2.
OS      Mus musculus (Mouse).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX      NCBI_TaxID=10090;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=BAB-14, and 129/SyJ;
RX      MEDLINE=2251062; PubMed=12016228;
RA      Teufel M., Fujimori M., Chashi Y., Higashi N., Onami T.M.,
RA      Hedrick S.M., Irimura T.;
RT      Molecular Cloning and Characterization of a Novel Mouse Macrophage C-
RT      type Lectin, mMG2, Which Has a Distinct Carbohydrate Specificity from
RT      mMG1.1;
RL      J. Biol. Chem. 277:28892-28901(2002).
DR      EMBL: AY103461; AA052097.1; -
DR      EMBL: AY103462; AA052098.1; -
DR      MGD; MGT2385729; MGL2.
DR      GO; GO:0005529; F: sugar binding; IEA.
DR      GO; GO:0007157; P: heterophilic cell adhesion; IEA.
DR      InterPro; IPR002353; AntiFreezeII.

```

```

DR      InterPro; IPR001304; Lectin C.
DR      InterPro; IPR005640; Lectin N.
DR      Pfam; PF00059; Lectin_C; 1.
DR      Pfam; PF03954; Lectin_N; 1.
DR      PRINTS; PR00356; ANTI-FREEZEII.
DR      SMART; SMC0034; CLECT; 1.
DR      PROSITE; PS00615; C-TYPE LECTIN 1; 1.
DR      PROSITE; PS00441; C-TYPE LECTIN 2; 1.
KW      Lectin.
SQ      SEQUENCE 332 AA; 38067 MW; 76167D0D55E253E2 CRC64;

Query Match      65.2%; Score 45; DB 11; Length 332;
Best Local Similarity 87.5%; Pred. No. 30;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      4 SYDGSWMD 11
      : ||||| :
DB      294 SYDGRWMD 301

RESULT 7
Q850Y4  PRELIMINARY; PRT; 342 AA.
AC      Q850Y4;
DT      01-JUN-2003 (TREMBlrel. 24, Created)
DT      01-JUN-2003 (TREMBlrel. 24, Last sequence update)
DT      01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE      Hypothetical protein (putative mutator-like transposase).
GN      OSUNB0027B08.21 OR OSUNB0078D06.6.
OS      Oryza sativa (Japonica cultivar-group).
OC      Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC      Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC      Ehrhacoidae; Oryzaceae; Oryza.
OX      NCBI_TaxID=33947;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=cv. Nipponbare;
RA      Buell C.R., Yuan Q., Ouyang S., Liu J., Gansberger K., Jones K.M.,
RA      Overton II L.L., Tsirlin T., Kim M.M., Bera U.J., Jin S.S.,
RA      Padrosh D.W., Tallon L.J., Koo H., Zisman V., Hsiao J., Blunt S.,
RA      Vanaken S.S., Riedmiller S.B., Utecherback T.T., Feldblyum T.V.,
RA      Yang Q.Q., Haas B.U., Suh B.B., Peterson J.J., Quackenbush J.,
RA      White O., Salzberg S.L., Fraser C.M.;
RT      "Oryza sativa chromosome 3 BAC OSUNB0027B08 genomic sequence.";
RL      Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.
RN      [2]
RP      SEQUENCE FROM N.A.
RC      STRAIN=cv. Nipponbare;
RA      Buell C.R., Yuan Q., Ouyang S., Liu J., Gansberger K., Jones K.M.,
RA      Overton II L.L., Tsirlin T., Kim M.M., Bera U.J., Jin S.S.,
RA      Padrosh D.W., Tallon L.J., Koo H., Zisman V., Hsiao J., Blunt S.,
RA      Vanaken S.S., Riedmiller S.B., Utecherback T.T., Feldblyum T.V.,
RA      Yang Q.Q., Haas B.U., Suh B.B., Peterson J.J., Quackenbush J.,
RA      White O., Salzberg S.L., Fraser C.M.;
RT      "Oryza sativa chromosome 3 BAC OSUNB0078D06 genomic sequence.";
RL      Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RN      [3]
RP      SEQUENCE FROM N.A.
RC      STRAIN=cv. Nipponbare;
RA      Buell R.;
RL      Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
DR      EMBL; AC133378; AA039856.1; -
DR      EMBL; AC133339; AAP46248.1; -
DR      InterPro; IPR004332; MUDR.
DR      Pfam; PF03108; MUDR; 1.
KW      Hypothetical protein.
SQ      SEQUENCE 342 AA; 39429 MW; B35548C32DC783AE CRC64;

Query Match      65.2%; Score 45; DB 10; Length 342;
Best Local Similarity 54.5%; Pred. No. 31;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY      2 ISSYDGSWMDY 12

```

Db 268 VHAYKGMNDY 278

## RESULT 8

Q8BUD5 PRELIMINARY; PRT; 361 AA.

AC Q8BUD5; 01-MAR-2003 (TREMBlrel. 23, Created)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)  
 DE 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
 DE Similar to macrophage galactose N-acetyl-galactosamine specific lectin.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=Mammary gland;  
 RX MEDLINE=22354683; PubMed=12466851;  
 RA The FANTOM Consortium,  
 RA The RIKEN Genome Exploration Research Group Phase I & II Team;  
 RT "Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs."  
 RL Nature 420:563-573(2002).  
 DR EMBL: AK085751; BAC39530.1; -  
 DR GO: 0005529; F-sugar binding; IEA.  
 DR InterPro: IPR01304; Lectin\_C.  
 DR InterPro: IPR005640; Lectin\_N.  
 DR Pfam: PF00059; lectin\_C; 1.  
 DR Pfam: PF03954; lectin\_N; 1.  
 DR SMART: SM00034; CLECT; 1.  
 DR PROSITE: PS00615; C-TYPE\_LECTIN\_2; 1.  
 DR PROSITE: PS50041; C-TYPE\_LECTIN\_2; 1.  
 SQ SEQUENCE 361 AA; 41010 MW; AACB8748C6A231 CRC64;

Query Match 65.2%; Score 45; DB 11; Length 361;  
 Best Local Similarity 87.5%; Pred. No. 33;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 SYDGSND 11

Db 323 SYDGRND 330

## RESULT 9

O81876 PRELIMINARY; PRT; 466 AA.

AC O81876; 01-NOV-1998 (TREMBlrel. 08, Created)  
 DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)  
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
 DE Hypothetical protein.  
 GN TIGL1.80 OR AT4G33590.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RA Obermaier B., Deutschenbaur S., Piravandi E., Hohnel J., Jesse T.,  
 RA Heijnen L., Vos P., Mewes H.W., Mayer K.F.X., Scheller C., Bevan M.,  
 RN Submitted (Aug-1998) to the EMBL/GenBank/DBJ databases.

RA SEQUENCE FROM N.A.

RA Obermaier B., Deutschenbaur S., Piravandi E., Mewes H.W., Lemcke K.,  
 RA Mayer K.F.X.;

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

RN [3]  
 RP SEQUENCE FROM N.A.

RA EU Arabidopsis sequencing project;

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL: AL031394; CA20572.1; -

DR EMBL: AL161583; CAB80076.1; -

DR PIR: T04976; T04976.

KW Hypothetical protein.

SQ SEQUENCE 466 AA; 53537 MW; AFA5334895A06858 CRC64;

Query Match 65.2%; Score 45; DB 10; Length 466;  
 Best Local Similarity 54.5%; Pred. No. 44;  
 Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 ISSYDGSNDY 12

Db 134 VDEHDSMSY 144

## RESULT 10

O81N69 PRELIMINARY; PRT; 655 AA.

AC O81N69; 01-OCT-2002 (TREMBlrel. 22, Created)  
 DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
 DE Mutator-like transposase, 3'-partial (Fragment).  
 GN OSJNB0091N21.31.  
 OS Oryza sativa (Japonica cultivar-group).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhacridae; Oryzaceae; Oryza.  
 NCBI\_TaxID=39947;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RC STRAIN=cv. Nipponbare;  
 RA Buell C.R., Yuan Q., Ouyang S., Liu J., Gansberger K., Kim M.M.,  
 RA Overton II L.L., Bera U.J., Taitin T., Krol M.I., Jarrari B.B.,  
 RA Jin S.S., Koo H., Zismann V., Hsiao J., Blunt S., Vanaken S.S.,  
 RA Utebäck T.T., Feldblyum T.V., Yang Q.Q., Haas B.J., Sub B.B.,  
 RA Peterson J.J., Quackenbush J., White O., Salzberg S., Fraser C.M.,  
 RT "Oryza sativa chromosome 10 BAC OSJNB0091N21 genomic sequence."  
 RL Submitted (Aug-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL: AC091122; AAM94925.1; -

DR Gramene; O81N69; -

DR InterPro: IPR004332; MDR.

DR Pfam: PF03108; MDR; 1.

FT NON\_TER 655

SQ SEQUENCE 655 AA; 76040 MW; EA18FED58B7FDBF0 CRC64;

Query Match 65.2%; Score 45; DB 10; Length 655;  
 Best Local Similarity 54.5%; Pred. No. 64;  
 Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 ISSYDGSNDY 12

Db 268 VHAYKGMNDY 278

## RESULT 11

Q7Y0C0 PRELIMINARY; PRT; 773 AA.

AC Q7Y0C0; 01-OCT-2003 (TREMBlrel. 25, Created)  
 DT 01-OCT-2003 (TREMBlrel. 25, Last sequence update)  
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
 DE Putative mutator-like transposase.  
 GN OSJNB007951.18.  
 OS Oryza sativa (Japonica cultivar-group).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhacridae; Oryzaceae; Oryza.  
 NCBI\_TaxID=39947;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RC STRAIN=cv. Nipponbare;

RA Buell C.R., Yuan Q., Ouyang S., Liu J., Gansberger K., Jones K.M.,  
 RA Overton II L.W., Taitlin T., Kim M.M., Bera J.J., Jin S.S.,  
 RA Padrosh D.W., Tallon L.J., Koo H., Zismann V., Hsiao J., Blunt S.,  
 RA Vanaken S.S., Riedmuller S.B., Utecherack T.T., Feldlym T.V.,  
 RA Yang Q.Q., Haas B.J., Suh B.B., Peterson J.J., Quackenbush J.,  
 RA White O., Salzberg S.L., Fraser C.M.,  
 RT "Oryza sativa chromosome 3 BAC OSJNB0079815 genomic sequence";  
 RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.  
 RN  
 RP  
 RC STRAIN=cv. Nipponbare;  
 RA Buell R.;  
 RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AC099043; AAF50964.1; -  
 SQ SEQUENCE 773 AA; 88785 MW; 7BA8C501103D8CB2 CRC64;

Query Match 65.2%; Score 45; DB 10; Length 773;  
 Best Local Similarity 54.5%; Pred. No. 77;  
 Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 ISSYDGSWMDY 12  
 Db 174 VHAYKGMNDY 184

## RESULT 12

08H8E2 PRELIMINARY; PRT; 779 AA.

AC 08H8E2;  
 DT 01-MAR-2003 (TREMBlrel. 23, Created)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)  
 DE 01-JUN-2003 (TREMBlrel. 24, Last annotation update)  
 DE Putative transposase protein.  
 GN OJ1006P06.3.  
 OS Oryza sativa (japonica cultivar-group).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoideae; Oryzaceae; Oryza.  
 OX NCBI\_TaxID=3947;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Wang R.A., Yu F., Soderlund C., Kim H.-R., Rambo T., Saski C.,  
 RA Currie J., Collura K.,  
 RT "Rice Genomic Sequence."  
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AC099399; AAN05493.1; -  
 DR InterPro: IPR004332; MUDR.  
 DR Pfam: PF03108; MUDR; 1.  
 SQ SEQUENCE 779 AA; 87872 MW; 6A9D0335FE26C955 CRC64;

Query Match 65.2%; Score 45; DB 10; Length 779;  
 Best Local Similarity 54.5%; Pred. No. 78;  
 Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 ISSYDGSWMDY 12  
 Db 268 VHAYKGMNDY 278

## RESULT 13

07XTH1 PRELIMINARY; PRT; 794 AA.

AC 07XTH1;  
 DT 01-OCT-2003 (TREMBlrel. 25, Created)  
 DT 01-OCT-2003 (TREMBlrel. 25, Last sequence update)  
 DE OSJNB0026104.5 protein.  
 GN OSJNB0026104.5.  
 OS Oryza sativa (Rice).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoideae; Oryzaceae; Oryza.  
 OX NCBI\_TaxID=4530;

RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Fu G., Wang S.Y., Ren S.X., Lv G., Lin W., Gu W.Q., Zhu G.F., Tu Y.F.,  
 RA Jia J., Sun H.F., Zhang Y., Cai Z., Chen J., Kang H., Chen X.Y.,  
 RA Shao Y., Yin Y., Hu Q.P., Zhang X.L., Zhang W., Wang L.J., Ding C.W.,  
 RA Sheng H.H., Gu J.L., Chen S.T., Ni L., Zhu F.H., Han B., Feng Q.,  
 RA Huang Y.C., Li Y., Zhu J.J., Zhao Q., Hu X., Liu Y.L., Xu Z.,  
 RA Chen L., Fan D.L., Weng Q.J., Zhang L., Lu Y.Q., Yu S.L., Liu X.H.,  
 RA Lu T.T., Zhang Y.J., Lu Y., Li C., Li T., Zhang Y., Hu H., Jia P.X.,  
 RA Qian Y.M., Ying K., Zhou B., Chen Z.H., Hao P., Zhang L., Wu M.,  
 RA Zhang R.Q., Guan J.P., Hong G.F.;  
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AL606443; CAB01500.1; -  
 SQ SEQUENCE 794 AA; 91076 MW; 90FD1473227871D5 CRC64;

Query Match 65.2%; Score 45; DB 10; Length 794;  
 Best Local Similarity 54.5%; Pred. No. 79;  
 Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 ISSYDGSWMDY 12  
 Db 268 VHAYKGMNDY 278

## RESULT 14

07XLO1 PRELIMINARY; PRT; 804 AA.

AC 07XLO1;  
 DT 01-OCT-2003 (TREMBlrel. 25, Created)  
 DT 01-OCT-2003 (TREMBlrel. 25, Last sequence update)  
 DE 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
 DE OSJNB0044M19.5 protein.  
 GN OSJNB0044M19.5.  
 OS Oryza sativa (Rice).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoideae; Oryzaceae; Oryza.  
 OX NCBI\_TaxID=4530;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Han B., Feng Q., Huang Y.C., Li Y., Zhu J.J., Zhao Q., Hu X.,  
 RA Liu Y.L., Mu J., Yu Z., Chen L., Fan D.L., Weng Q.J., Zhang L.,  
 RA Lu Y.Q., Yu S.L., Liu X.H., Lu T.T., Zhang Y.J., Lu Y., Li C., Li T.,  
 RA Zhang Y., Hu H., Jia P.X., Qian Y.M., Ying K., Zhou B., Chen Z.H.,  
 RA Hao P., Zhang L., Wu M., Zhang R.Q., Guan J.P., Fu G., Wang S.Y.,  
 RA Ren S.X., Lv G., Lin W., Gu W.Q., Zhu G.F., Tu Y.F., Jia J., Yin H.F.,  
 RA Zhang Y., Cai Z., Chen J., Kang H., Chen X.Y., Shao C.Y., Sun Y.,  
 RA Hu Q.P., Zhang X.L., Zhang W., Wang L.J., Ding C.W., Sheng H.H.,  
 RA Gu J.L., Chen S.T., Ni L., Zhu F.H., Hong G.F.;  
 RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AL71601; CAB05018.1; -  
 SQ SEQUENCE 804 AA; 91227 MW; 172D5556EC7B7327 CRC64;

Query Match 65.2%; Score 45; DB 10; Length 804;  
 Best Local Similarity 54.5%; Pred. No. 80;  
 Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 ISSYDGSWMDY 12  
 Db 162 VHAYKGMNDY 172

## RESULT 15

084R42 PRELIMINARY; PRT; 844 AA.

AC 084R42;  
 DT 01-JUN-2003 (TREMBlrel. 24, Created)  
 DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)  
 DE 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
 DE Putative mutator-like transposase.  
 GN OSJNB0016H12.24.  
 OS Oryza sativa (japonica cultivar-group).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

```

OC  Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC  Ehrhartoideae; Oryzeae; Oryza.
OX  NCBI_TaxID=39947;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=cv. Nipponbare;
RA  Buehl C.R., Yuan Q., Ouyang S., Liu J., Gansberger K., Jones K.M.,
RA  Overton II L.L., Taitzin T., Kim M.M., Bera J.J., Jin S.S.,
RA  Padrosh D.W., Tallon L.J., Koo H., Zismann V., Hsiao J., Blunt S.,
RA  Vanaken S.S., Riedmuller S.B., Uteback T.T., Feldlyum T.V.,
RA  Yang Q.Q., Haas B.J., Suh B.B., Peterson J.J., Quackenbush J.,
RA  White O., Salzberg S.L., Fraser C.M.;
RT  "Oryza sativa chromosome 3 BAC OSUNB0016H12 genomic sequence.";
RL  Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
RN  [2]
RP  SEQUENCE FROM N.A.
RC  STRAIN=cv. Nipponbare;
RA  Buehl R.;
RL  Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
DR  EMBL; AC118133; AAP03412.1;
DR  GO; GO:0003676; F:nucleic acid binding; IEA.
DR  InterPro; IPR004332; MDR.
DR  InterPro; IPR001878; ZnF_CCHC.
DR  InterPro; IPR006564; ZnF_PMZ.
DR  InterPro; IPR007527; ZnF_SWIM.
DR  Pfam; PF03108; MDR; 1.
DR  Pfam; PF04434; SWIM; 1.
DR  PRINTS; PR00339; C2HCZNFINGER.
DR  SMART; SM00575; ZnF_PMZ; 1.
SQ  SEQUENCE 844 AA; 95790 MW; C011F7F209F3426 CRC64;

```

Query Match 65.2%; Score 45; DB 10; Length 844;

Best Local Similarity 54.5%; Pred. NO. 85;

Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Oy 2 ISSYDGSWNDY 12

DB 242 VHAYKGKWNXY 252

Search completed: October 5, 2004, 08:13:34  
Job time : 9.35185 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2004, 07:04:10 ; Search time 9.12963 Seconds  
(without alignments)  
371.381 Million cell updates/sec

Title: US-09-805-290A-15  
Perfect score: 69  
Sequence: 1 LISSYDGSWMDY 12

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_23Jan04:\*  
1: Geneseqp1980s:\*  
2: Geneseqp1990s:\*  
3: Geneseqp2000s:\*  
4: Geneseqp2001s:\*  
5: Geneseqp2002s:\*  
6: Geneseqp2003as:\*  
7: Geneseqp2003bs:\*  
8: Geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	69	100.0	12	AAE10548	Llama spe
2	69	100.0	129	AAE10558	AAE10558 HPL inhib
3	47	68.1	324	ADB74513	ADB74513 Mycobacte
4	45	63.2	1647	ABW67419	ABW67419 Photornab
5	44	63.8	12	AAE10547	AAE10547 Llama spe
6	44	63.8	129	AAE10551	AAE10551 HPL inhib
7	42	60.9	88	AAU63004	AAU63004 Propionib
8	42	60.9	88	ABM59523	ABM59523 Propionib
9	42	60.9	385	ABP66065	ABP66065 Bifidobac
10	42	60.9	543	ABBS58674	ABBS58674 Drosophil
11	42	60.9	545	ABBS58250	ABBS58250 Drosophil
12	42	60.9	556	ABBS67284	ABBS67284 Drosophil
13	41	59.4	295	ADBS63678	ADBS63678 Rat Prote
14	41	59.4	295	ADBS63680	ADBS63680 Rat Prote
15	41	59.4	297	AAE10547	AAE10547 M. tuberc
16	41	59.4	297	ADA10972	ADA10972 Human CDN
17	41	59.4	382	AAE82272	AAE82272 Thiosulph
18	41	59.4	576	AAE78613	AAE78613 Expressio
19	41	59.4	592	AAE92527	AAE92527 Fas anticg
20	41	59.4	596	AAE78616	AAE78616 Expressio
21	41	59.4	600	AAE78610	AAE78610 Expressio
22	41	59.4	600	AAE92526	AAE92526 Fas anticg
23	41	59.4	878	AAE78608	AAE78608 Murine IL
24	41	59.4	878	AAE92529	AAE92529 Fas seque
25	41	59.4	1046	AAW02156	AAW02156 Periplasm

26	41	59.4	1046	AAE52304	AAE52304 Vibrio fu
27	40	58.0	111	ABR55874	ABR55874 Human mAb
28	40	58.0	111	ABR55875	ABR55875 Human mAb
29	40	58.0	157	AAE6115	AAE6115 S. pneumo
30	40	58.0	237	AAE21394	AAE21394 Arabidops
31	40	58.0	237	AAE21394	AAE21394 Arabidops
32	40	58.0	237	AAE21394	AAE21394 Arabidops
33	40	58.0	237	AAE21394	AAE21394 Arabidops
34	40	58.0	237	AAE21394	AAE21394 Arabidops
35	40	58.0	237	AAE21394	AAE21394 Arabidops
36	40	58.0	237	AAE21394	AAE21394 Arabidops
37	40	58.0	237	AAE21394	AAE21394 Arabidops
38	40	58.0	237	AAE21394	AAE21394 Arabidops
39	40	58.0	237	AAE21394	AAE21394 Arabidops
40	40	58.0	237	AAE21394	AAE21394 Arabidops
41	40	58.0	237	AAE21394	AAE21394 Arabidops
42	40	58.0	237	AAE21394	AAE21394 Arabidops
43	40	58.0	237	AAE21394	AAE21394 Arabidops
44	40	58.0	237	AAE21394	AAE21394 Arabidops
45	40	58.0	237	AAE21394	AAE21394 Arabidops

## ALIGNMENTS

RESULT 1  
AAE10548  
ID AAE10548 standard; peptide; 12 AA.  
XX  
AC AAE10548;  
XX  
XX  
DT 10-DEC-2001 (first entry)  
XX  
DE Llama species antibody VHH CDR3 #17.  
XX  
KW Llama antibody; camelid; anorectic; heavy chain variable domain; VHH;  
KW human dietary enzyme inhibitor; medicament; human pancreatic lipase; HPL;  
KW food; human gastric lipase; HGL; cosmetic control; body weight;  
KW complementarity determining region 3; CDR3.  
XX  
OS Llama sp.  
XX  
PN EP1134231-A1.  
XX  
PD 19-SEP-2001.  
XX  
PF 20-FEB-2001; 2001EP-00200703.  
XX  
PR 14-MAR-2000; 2000EP-00200930.  
XX  
PA (UNITL ) UNILEVER NV.  
PA (UNITL ) UNILEVER PLC.  
XX  
PI Bezemer S, Van De Burg M, De Haard JWM, Tareilus B;  
XX WPI; 2001-572718/65.  
XX  
DR New antibody or its fragments for inhibiting human dietary enzymes,  
XX useful for cosmetic control of body weight of human beings, comprises  
XX heavy chain variable domain derived from immunoglobulin naturally devoid  
XX of light chains.  
XX  
PS Claim 4; Page 29; 37pp; English.  
XX  
CC The patent discloses antibodies or their fragments comprising a heavy  
CC chain variable domain (VHH) derived from an immunoglobulin naturally  
CC devoid of light chains specific for inhibiting human dietary enzymes. The  
CC antibodies of the invention are useful for the preparation of medicaments  
CC or food for inhibiting the activity of one or more human dietary enzymes  
CC especially human pancreatic lipase (HPL) or human gastric lipase (HGL)  
CC which are useful for the cosmetic control of body weight of human beings.  
CC The present peptide sequence is a complementarity determining region 3  
CC (CDR3) of llama species (camelid) antibody VHH region

```

XX      SQ      Sequence 12 AA;
XX
XX      Query Match      100.0%;      Score 69;      DB 4;      Length 12;
XX      Best Local Similarity      100.0%;      Freq. No. 0.00061;
XX      Matches      12;      Conservative      0;      Mismatches      0;      Indels      0;      Gaps      0
XX
XX      QY      1      LISSYDGSWNDY      12
XX      1      |||||
XX      Db      1      LISSYDGSWNDY      12
XX
XX      RESULT 2
XX      ID      AAE10558
XX      AAE10558      standard; peptide; 129 AA.
XX
XX      AC      AAE10558;
XX      XX
XX      DT      10-DEC-2001      (first entry)
XX      DE      HPL inhibiting VHH fragment, HPL #22 from llama species.
XX      XX
XX      KW      Llama antibody; camelid; anorectic; heavy chain variable domain; VHH;
XX      human dietary enzyme inhibitor; medicament; human pancreatic lipase; HPL;
XX      food; human gastric lipase; HGL; cosmetic control; body weight.
XX      XX
XX      OS      Lama sp.
XX      XX
XX      FH      Key      Location/Qualifiers
XX      FH      Region      31..35
XX      FT      /label= CDR1
XX      FT      /note= "Complementarity determining region 1"
XX      FT      Region      50..64
XX      FT      /label= CDR2
XX      FT      /note= "Complementarity determining region 2"
XX      FT      Region      98..109
XX      FT      /label= CDR3
XX      FT      /note= "Complementarity determining region 3"
XX      FN      EPI134231-AA1.
XX      PD      19-SEP-2001.
XX      PF      20-FEB-2001; 2001EP-00200703.
XX      PR      14-MAR-2000; 2000EP-00200930.
XX      PA      (UNITIL ) UNILEVER NV.
XX      PA      (UNITIL ) UNILEVER PLC.
XX      PI      Bezemer S, Van De Burg M, De Haard JW, Tareilus E,
XX      DR      WPI; 2001-572718/65.
XX      XX
XX      PT      New antibody or its fragments for inhibiting human dietary enzymes,
XX      PT      useful for cosmetic control of body weight of human beings, comprises
XX      PT      heavy chain variable domain derived from immunoglobulin naturally devoid
XX      PT      of light chains.
XX      PS      Example 2; Page 10; 37pp; English.
XX      XX
XX      CC      The patent discloses antibodies or their fragments comprising a heavy
XX      CC      chain variable domain (VHH) derived from an immunoglobulin naturally
XX      CC      devoid of light chains specific for inhibiting human dietary enzymes. The
XX      CC      antibodies of the invention are useful for the preparation of medicaments
XX      CC      or food for inhibiting the activity of one or more human dietary enzymes
XX      CC      especially human pancreatic lipase (HPL) or human gastric lipase (HGL)
XX      CC      which are useful for the cosmetic control of body weight of human beings.
XX      CC      The present peptide sequence is HPL inhibiting VHH fragment, HPL #22 from
XX      CC      llama (camelid) species
XX      XX
XX      Sequence 129 AA;
XX

```

[illegible]



```

XX Key Location/Qualifiers
FH Region 31..35
FT /label= CDR1
FT /note= "Complementarity determining region 1"
FT Region 50..64
FT /label= CDR2
FT /note= "Complementarity determining region 2"
FT Region 98..109
FT /label= CDR3
FT /note= "Complementarity determining region 3"
XX EPI134231-A1.
XX 19-SEP-2001.
XX 20-FEB-2001; 2001EP-00200703.
XX 14-MAR-2000; 2000EP-00200930.
XX (UNITL ) UNILEVER NV.
XX (UNITL ) UNILEVER PLC.
XX Bezemer S, Van De Burg M, De Haard JWM, Tareilus E;
XX WPI; 2001-572718/65.
XX DR
XX PT New antibody or its fragments for inhibiting human dietary enzymes,
XX PT useful for cosmetic control of body weight of human beings, comprises
XX PT heavy chain variable domain derived from immunoglobulin naturally devoid
XX PT of light chains.
XX PS
XX PS Example 2; Page 9; 37pp; English.
XX CC The patent discloses antibodies or their fragments comprising a heavy
XX CC chain variable domain (VH) derived from an immunoglobulin naturally
XX CC devoid of light chains specific for inhibiting human dietary enzymes. The
XX CC antibodies of the invention are useful for the preparation of medicaments
XX CC or food for inhibiting the activity of one or more human dietary enzymes
XX CC especially human pancreatic lipase (HPL) or human gastric lipase (HGL)
XX CC which are useful for the cosmetic control of body weight of human beings.
XX CC The present peptide sequence is HPL inhibiting VH fragment, HPL #11 from
XX CC Lama (camelid) species
XX SQ
XX SQ Sequence 129 AA;
XX Query Match 63.8%; Score 44; DB 4; Length 129;
XX Best Local Similarity 70.0%; Pred. No. 49;
XX Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 3 SSYDGSWMDY 12
Db 100 SNYDRSMGDY 109

```

RESULT 7

AU63004 ID AU63004 standard; protein; 88 AA.

XX AC AAU63004;

XX DT 27-FEB-2002 (first entry)

DE Propionibacterium acnes immunogenic protein #23900.

XX SAHO syndrome; synovitis; acne; pustulosis; hypertosia; osteomyelitis;

KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;

KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;

KW dermatological; osteopathic; neuroprotectant.

XX Propionibacterium acnes.

OS W0200181581-A2.

XX PN

```

XX PD 01-NOV-2001.
XX 20-APR-2001; 2001WO-US012865.
XX 21-APR-2000; 2000US-0199047P.
XX 02-JUN-2000; 2000US-0208841P.
XX 07-JUL-2000; 2000US-0216747P.
XX (CORI ) CORIXA CORP.
XX Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;
XX L'laismonneuve J, Zhang Y, Jen S, Carter D;
XX WPI; 2001-616774/71.
XX DR N-PSDB; AAS59630.
XX PT Propionibacterium acnes polypeptides and nucleic acids useful for
XX PT vaccinating against and diagnosing infections, especially useful for
XX PT treating acne vulgaris.
XX PS
XX PS Example 1; SEQ ID NO 24199; 1069pp; English.
XX CC Sequences AU63005-AU68017 represent Propionibacterium acnes immunogenic
XX CC polypeptides. The proteins and their associated DNA sequences are used in
XX CC the treatment, prevention and diagnosis of medical conditions caused by
XX CC P. acnes. The disorders include SAHO syndrome (synovitis, acne,
XX CC pustulosis, hypertosia and osteomyelitis), uveitis and endophthalmitis.
XX CC P. acnes is also involved in infections of bone, joints and the central
XX CC nervous system, however it is particularly involved in the inflammatory
XX CC lesions associated with acne vulgaris. A method for detecting the
XX CC presence or absence of P. acnes in a patient comprises contacting a
XX CC sample with a binding agent that binds to the proteins of the invention
XX CC and determining the amount of bound protein in the sample. The
XX CC polypeptides may be used as antigens in the production of antibodies
XX CC specific for P. acnes proteins. These antibodies can be used to
XX CC downregulate expression and activity of P. acnes polypeptides and
XX CC therefore treat P. acnes infections. The antibodies may also be used as
XX CC diagnostic agents for determining P. acnes presence, for example, by
XX CC enzyme linked immunosorbent assay (ELISA). Note: The sequence data for
XX CC this patent did not form part of the printed specification, but was
XX CC obtained in electronic format directly from WIPO at
XX CC ftp.wipo.int/pub/published_ptc_sequences
XX SQ
XX SQ Sequence 88 AA;
XX Query Match 60.9%; Score 42; DB 4; Length 88;
XX Best Local Similarity 63.6%; Pred. No. 64;
XX Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
QY 1 LISSYDGSWMD 11
Db 14 VISTEDGSWMD 24

```

RESULT 8

ABM59523 ID ABM59523 standard; protein; 88 AA.

XX AC ABM59523;

XX DT 20-OCT-2003 (first entry)

DE Propionibacterium acnes predicted ORF-encoded polypeptide #24199.

XX Acne vulgaris; antiseborrheic; dermatological; antibacterial;

KW immunostimulant; immune response; vaccine.

XX Propionibacterium acnes.

OS W02003033515-A1.

XX PN

XX PD 24-APR-2003.

XX 11-OCT-2002; 2002WO-US032727.  
 PF  
 XX  
 PR 15-OCT-2001; 2001US-00978825.  
 XX  
 PA (CORI-) CORIXA CORP.  
 XX  
 PI Mitcham JL, Skeiky YAM, Persing DH, Bhatia A, Maisonneuve JJ,  
 PI Zhang Y, Wang S, Jen S, Lodes MJ, Benson DR, Jones R, Carter D;  
 PI Barth B, Vallieve-Douglas J;  
 XX  
 DR WPI, 2003-381789/36.  
 DR N-PSDB; ACF64559.  
 XX  
 PT New Propionibacterium acnes polypeptides and polynucleotides encoding the  
 PT polypeptide, useful for diagnosing, preventing or treating acne vulgaris,  
 PT or for stimulating an immune response specific for a P. acnes protein.  
 XX  
 PS Example 1; SEQ ID NO 24199; 1481bp; English.  
 XX  
 CC The invention relates to an isolated polynucleotide (ACF64435-ACF64733)  
 CC encoding a Propionibacterium acnes protein. The invention also relates to  
 CC polypeptides encoded by the polynucleotides (ABM35624-ABM64536) and to  
 CC immunogenic fragments of P. acnes polypeptides. The invention  
 CC additionally encompasses expression vectors and host cells comprising a  
 CC polynucleotide of the invention; antibodies against polypeptides of the  
 CC invention; fusion proteins comprising a polypeptide of the invention; a  
 CC method for stimulating an immune response specific for a P. acnes  
 CC polypeptide and an isolated T cell population comprising T cells prepared  
 CC via this method; a vaccine composition (comprising P. acnes polypeptides,  
 CC polynucleotides, antibodies, fusion proteins, T cell populations, or  
 CC antigen-presenting cells that express the polypeptide); a method and kit  
 CC for detecting or determining the presence or absence of P. acnes in a  
 CC patient; and a method for inhibiting the development of P. acnes in a  
 CC patient. The P. acnes polypeptides, polynucleotides, antibodies, fusion  
 CC proteins, T cell populations or antigen-presenting cells that express the  
 CC polypeptides are useful for diagnosing, preventing or treating acne  
 CC vulgaris, or for stimulating an immune response specific for a P. acnes  
 CC protein. The polynucleotides can also be used as probes or primers for  
 CC nucleic acid hybridization. The vaccine composition is useful for the  
 CC stimulation of an immune response against P. acnes, or for treating acne,  
 CC and the kit is useful for performing a diagnostic assay. The present  
 CC sequence represents a polypeptide predicted to be encoded by an ORF (open  
 CC reading frame) contained within the P. acnes polynucleotides of the  
 CC invention. Note: The sequence data for this patent did not form part of  
 CC the printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 88 AA:  
 Query Match 60.9%; Score 42; DB 6; Length 88;  
 Best Local Similarity 63.6%; Pred. No. 64;  
 Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 ISSYDGSWMD 11  
 :|||:||||:  
 Db 14 VISTEDGSWSD 24  
 RESULT 9  
 ABP66065  
 ID ABP66065 standard; protein; 385 AA.  
 XX  
 AC ABP66065;  
 XX  
 DT 19-NOV-2002 (first entry)  
 XX  
 DE Bifidobacterium longum NCC2705 ORF amino acid sequence SEQ ID NO:809.  
 XX  
 KW Bifidobacterium longum NCC2705; Bifidobacterium; bacterial;  
 KW anti-diarrheic; antibacterial; inhibitor of Salmonella; detection;  
 KW identification; lactic acid bacterium; diarrhoea; pathogenic bacteria;  
 KW rotavirus; food composition; pharmaceutical composition.

XX Bifidobacterium longum.  
 OS  
 XX  
 XX EPI227152-A1.  
 FN  
 XX  
 XX 31-JUL-2002.  
 PD  
 XX  
 XX 30-JAN-2001; 2001EP-00102050.  
 PF  
 XX  
 XX 30-JAN-2001; 2001EP-00102050.  
 PR  
 XX  
 XX (NEST ) SOC PROD NESTLE SA.  
 PA  
 DR WPI, 2002-668397/72.  
 DR  
 XX  
 XX  
 PT Novel polynucleotide comprising Bifidobacterium genome sequence useful as  
 PT a probe or primer for detecting and/or identifying Bifidobacterium longum  
 PT in a biological sample.  
 XX  
 PS Claim 3; SEQ ID NO 809; 80bp; English.  
 XX  
 CC The present invention describes a polynucleotide (I) comprising a  
 CC sequence of a Bifidobacterium genome selected from the nucleotide  
 CC sequences given in ABQ81842 and ABQ81843, or a sequence exhibiting at  
 CC least 90% identity or which hybridises with the sequences given in  
 CC ABQ81842 and ABQ81843. Also described is a polynucleotide (II) encoding a  
 CC fusion protein, comprising a sequence selected from 1097 sequences given  
 CC in ABP655258 to ABP66354 ligated in frame to a polynucleotide encoding a  
 CC heterologous polypeptide. (I) has anti-diarrheic and antibacterial  
 CC activities, and can be used as an inhibitor of Salmonella. (I) (which is  
 CC a probe) is useful for the detection and/or identification of  
 CC Bifidobacterium longum in a biological sample. A carrier containing the  
 CC lactic acid bacterium Bifidobacterium longum NCC2705 (NCIM I-2618) can be  
 CC used for preventing and/or treating diarrhoea brought about by pathogenic  
 CC bacteria and/or rotavirus. The carrier is a food composition selected  
 CC from milk, yogurt, curd, cheese, fermented milks, milk based fermented  
 CC products, ice-creams, fermented cereal based products, milk based  
 CC powders, infant formula, pet food or a pharmaceutical composition  
 CC selected from tablets, liquid bacterial suspensions, dried oral  
 CC supplement, wet oral supplement, dry tube feeding or wet tube feeding.  
 CC (I) is useful in DNA arrays or chips to carry out analysis of the  
 CC expression of the Bifidobacterium gene. ABQ81844 to ABQ81850 represent  
 CC Bifidobacterium related nucleotide sequences given in the sequence  
 CC listing from the present invention but not mentioned further within the  
 CC specification. N.B. The sequence data for this patent is not represented  
 CC in the printed specification but is based on sequence information  
 CC supplied by the European Patent Office  
 XX  
 SQ Sequence 385 AA:  
 Query Match 60.9%; Score 42; DB 5; Length 385;  
 Best Local Similarity 63.6%; Pred. No. 3,2e+02;  
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
 QY 2 ISSYDGSWMDY 12  
 :|||:||||:  
 Db 309 IFSFGSWMDY 319  
 RESULT 10  
 ABB58674  
 ID ABB58674 standard; protein; 543 AA.  
 XX  
 AC ABB58674;  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE Drosophila melanogaster polypeptide SEQ ID NO 2814.  
 XX  
 KW Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical.  
 XX  
 OS Drosophila melanogaster.

XX WO200171042-A2.  
 XX PD 27-SEP-2001.  
 XX PF 23-MAR-2001; 2001WO-US009231.  
 XX PR 23-MAR-2000; 2000US-0191637P.  
 XX PR 11-JUL-2000; 2000US-00614150.  
 XX PA (PEKE ) PE CORP NY.  
 XX PI Venter JC, Adams M, Li PWD, Myers EW;  
 XX DR WPI; 2001-656860/75.  
 XX DR N-PSDB; ABL02777.  
 XX PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signaling and cell-cell  
 PT interactions.  
 XX PS Disclosure; SEQ ID NO 2814; 21pp + Sequence Listing; English.  
 XX CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signaling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 CC sequences (ABL01840-ABL16175) and the encoded proteins (ABBS7737-  
 CC ABB72072). The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
 XX SQ Sequence 543 AA;  
 XX  
 XX Query Match 60.9%; Score 42; DB 4; Length 543;  
 XX Best Local Similarity 77.8%; Pred. No. 4.7e+02;  
 XX Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 2 ISSYDGSWN 10  
 Db 84 ISKXDGIMN 92  
 XX  
 XX RESULT 11  
 XX ABB58250  
 XX ID ABB58250 standard; protein; 545 AA.  
 XX AC ABB58250;  
 XX XX  
 XX DT 26-MAR-2002 (first entry)  
 XX DE Drosophila melanogaster polypeptide SEQ ID NO 1542.  
 XX KW Drosophila; developmental biology; cell signalling; insecticide;  
 XX KW pharmaceutical.  
 XX OS Drosophila melanogaster.  
 XX PI Venter JC, Adams M, Li PWD, Myers EW;  
 XX DR WPI; 2001-656860/75.  
 XX DR N-PSDB; ABL11387.  
 XX PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signaling and cell-cell  
 PT interactions.  
 XX PS Disclosure; SEQ ID NO 1542; 21pp + Sequence Listing; English.  
 XX CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signaling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention

DR WPI; 2001-656860/75.  
 DR N-PSDB; ABL02353.  
 XX  
 XX PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signaling and cell-cell  
 PT interactions.  
 XX PS Disclosure; SEQ ID NO 1542; 21pp + Sequence Listing; English.  
 XX CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signaling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 CC sequences (ABL01840-ABL16175) and the encoded proteins (ABBS7737-  
 CC ABB72072). The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
 XX SQ Sequence 545 AA;  
 XX  
 XX Query Match 60.9%; Score 42; DB 4; Length 545;  
 XX Best Local Similarity 77.8%; Pred. No. 4.7e+02;  
 XX Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 2 ISSYDGSWN 10  
 Db 42 ISKXDGIMN 50  
 XX  
 XX RESULT 12  
 XX ABB67284  
 XX ID ABB67284 standard; protein; 556 AA.  
 XX AC ABB67284;  
 XX XX  
 XX DT 26-MAR-2002 (first entry)  
 XX DE Drosophila melanogaster polypeptide SEQ ID NO 28644.  
 XX KW Drosophila; developmental biology; cell signalling; insecticide;  
 XX KW pharmaceutical.  
 XX OS Drosophila melanogaster.  
 XX PI Venter JC, Adams M, Li PWD, Myers EW;  
 XX DR WPI; 2001-656860/75.  
 XX DR N-PSDB; ABL11387.  
 XX PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signaling and cell-cell  
 PT interactions.  
 XX PS Disclosure; SEQ ID NO 28644; 21pp + Sequence Listing; English.  
 XX CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signaling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention

CC discloses genomic DNA sequences (AB16176-AB163511), expressed DNA  
CC sequences (AB161840-AB16175) and the encoded proteins (AB57737-  
CC AB572072). The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pat\_sequences  
CC XX

Sequence 556 AA;

Query Match 60.9%; Score 42; DB 4; Length 556;

Best Local Similarity 77.8%; Pred. No. 4.8e+02; Indels 0; Gaps 0;  
Matches 7; Conservative 0; Mismatches 2;

QY 2 ISSYDGSNN 10  
DB 84 ISKYGDIWN 92

RESULT 13  
ADE63678  
ID ADE63678 standard; protein; 295 AA.

AC ADE63678;

DT 29-JAN-2004 (first entry)

DE Rat Protein P24329, SEQ ID NO 9622.

XX Rat; pain; neuronal tissue; gene therapy; spinal segmental nerve injury;  
XX chronic constriction injury; CCI; spared nerve injury; SNI; Chung.

OS Rattus norvegicus.

XX MO2003016475-A2.

XX 27-FEB-2003.

PD 14-AUG-2002; 2002MO-US025765.

XX 14-AUG-2001; 2001US-0312147P.

PR 01-NOV-2001; 2001US-0346382P.

PR 26-NOV-2001; 2001US-0333347P.

XX (GENO) GEN HOSPITAL CORP.

PA (FARB) BAYER AG.

PI Woolf C, D'urso D, Befort K, Costigan M,

DR WPI; 2003-268312/26.

XX GENBANK; P24329.

PT New composition comprising two or more isolated polypeptides, useful for

PT preparing a medicament for treating pain in an animal.

XX Claim 1; Page; 1017pp; English.

CC The invention discloses a composition comprising two or more isolated rat  
CC or human polynucleotides or a polynucleotide which represents a fragment,  
CC derivative or allelic variation of the nucleic acid sequence. Also  
CC claimed are a vector comprising the novel polynucleotide, a host cell  
CC comprising the vector, a method for identifying a nucleotide sequence  
CC which is differentially regulated in an animal subjected to pain and a  
CC kit to perform the method, an array, a method for identifying an agent  
CC that increases or decreases the expression of the polynucleotide sequence  
CC that is differentially expressed in neuronal tissue of a first animal  
CC subjected to pain, a method for identifying a compound which regulates  
CC the expression of a polynucleotide sequence which is differentially  
CC expressed in an animal subjected to pain, a method for identifying a  
CC compound that regulates the activity of one or more of the  
CC polynucleotides, a method for producing a pharmaceutical composition, a  
CC method for identifying a compound or small molecule that regulates the  
CC activity in an animal of one or more of the polypeptides given in the  
CC specification, a method for identifying a compound useful in treating  
CC pain and a pharmaceutical composition comprising the one or more

CC polypeptides or their antibodies. The polynucleotide or the compound that  
CC modulates its activity is useful for preparing a medicament for treating  
CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction  
CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene  
CC therapy). The sequence presented is a rat protein (shown in Table 2 of  
CC the specification) which is differentially expressed during pain. Note:  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic form directly from WIPO at  
CC ftp.wipo.int/pub/published\_pat\_sequences.

Sequence 295 AA;

Query Match 59.4%; Score 41; DB 7; Length 295;

Best Local Similarity 45.5%; Pred. No. 3.4e+02; Indels 0; Gaps 0;  
Matches 5; Conservative 5; Mismatches 1;

QY 2 ISSYDGSNN 12  
DB 267 VAVYDGSNN 277

RESULT 14  
ADE63680  
ID ADE63680 standard; protein; 295 AA.

AC ADE63680;

DT 29-JAN-2004 (first entry)

DE Rat Protein P24329, SEQ ID NO 9624.

XX Rat; pain; neuronal tissue; gene therapy; spinal segmental nerve injury;  
XX chronic constriction injury; CCI; spared nerve injury; SNI; Chung.

OS Rattus norvegicus.

XX MO2003016475-A2.

XX 27-FEB-2003.

PD 14-AUG-2002; 2002MO-US025765.

XX 14-AUG-2001; 2001US-0312147P.

PR 01-NOV-2001; 2001US-0346382P.

PR 26-NOV-2001; 2001US-0333347P.

XX (GENO) GEN HOSPITAL CORP.

PA (FARB) BAYER AG.

PI Woolf C, D'urso D, Befort K, Costigan M,

DR WPI; 2003-268312/26.

XX GENBANK; P24329.

PT New composition comprising two or more isolated polypeptides, useful for

PT preparing a medicament for treating pain in an animal.

XX Claim 1; Page; 1017pp; English.

CC The invention discloses a composition comprising two or more isolated rat  
CC or human polynucleotides or a polynucleotide which represents a fragment,  
CC derivative or allelic variation of the nucleic acid sequence. Also  
CC claimed are a vector comprising the novel polynucleotide, a host cell  
CC comprising the vector, a method for identifying a nucleotide sequence  
CC which is differentially regulated in an animal subjected to pain and a  
CC kit to perform the method, an array, a method for identifying an agent  
CC that increases or decreases the expression of the polynucleotide sequence  
CC that is differentially expressed in neuronal tissue of a first animal  
CC subjected to pain, a method for identifying a compound which regulates  
CC the expression of a polynucleotide sequence which is differentially  
CC expressed in an animal subjected to pain, a method for identifying a  
CC compound that regulates the activity of one or more of the  
CC polynucleotides, a method for producing a pharmaceutical composition, a

CC method for identifying a compound or small molecule that regulates the  
 CC activity in an animal of one or more of the polypeptides given in the  
 CC specification, a method for identifying a compound useful in treating  
 CC pain and a pharmaceutical composition comprising the one or more  
 CC polypeptides or their antibodies. The polynucleotide or the compound that  
 CC modulates its activity is useful for preparing a medicament for treating  
 CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction  
 CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene  
 CC therapy). The sequence presented is a rat protein (shown in Table 2 of  
 CC the specification) which is differentially expressed during pain. Note:  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic form directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.

CC Sequence 295 AA;

Query Match 59.4%; Score 41; DB 7; Length 295;

Best Local Similarity 45.5%; Pred. No. 3.4e+02; Mismatches 1; Indels 0; Gaps 0;

Matches 5; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 2 ISSYDGSWMDY 12

DB 267 VAVYDGSWSEW 277

RESULT 15

AA87871 ID AA87871 standard; protein; 297 AA.

AA87871; AC

06-OCT-2000 (first entry) DE

M. tuberculosis antigen TB33 protein. DE

Tuberculosis; TB; antigen; vaccine; diagnosis; somatic; tuberculostatic; KM

infection; interferon-gamma; IFN-gamma; protective immunity; therapy; KW

delayed type hypersensitivity response; TB33. XX

Mycobacterium tuberculosis. OS

WO200021983-A2. PN

20-APR-2000. PD

08-OCT-1999; 99WO-DK000538. PF

08-OCT-1998; 98DK-00001281. PR

21-JAN-1999; 99US-0116673P. PR

(STAT-) STATENS SERUM INST. PA

Andersen P, Weidinger K, Hansen CV, Florio W, Okkels LMW; PI

Skjot RLV, Rosenkrands I; DR

WPI; 2000-317931/27. DR

N-PSDB; AAA39565. DR

Novel Polypeptide of somatic protein extract useful as vaccine against PT

virulent Mycobacterium infection, isolated from cell wall, cell membrane PT

and cytosol. PS

Claim 1; Page 91; 126pp; English. XX

This invention describes a novel polypeptide (PP) of somatic proteins  
 CC extract (I) which have tuberculostatic activity. (I) or their subsequence  
 CC has at least one of the following properties: (a) the PP induces an in  
 CC vitro recall response, or an in vitro response, during primary infection  
 CC with virulent Mycobacterium, determined by a release of interferon (IFN)-  
 CC gamma, (b) PP induces a protective immunity, determined by vaccinating an  
 CC animal with PP and an adjuvant, three times at two weeks intervals, (c)  
 CC PP induces an in vitro response, or in vitro recall response, determined  
 CC by release of IFN-gamma of at least 1000 pg/ml or 500 pg/ml,

CC respectively, from peripheral Blood Mononuclear Cells (PBMC) withdrawn  
 CC from TB patients, or PPD positive individuals, 6 months after diagnosis,  
 CC (d) PP induces a specific antibody response in a TB patient, as  
 CC determined by enzyme linked immunosorbent assay (ELISA) technique or a  
 CC western blot, (e) PP induces a positive delayed type hypersensitivity  
 CC (DTH) response, determined by intradermal injection. (I) and (II) are  
 CC useful in preparing a prophylactic or therapeutic medicine as a vaccine  
 CC for induction of a protective or generation of an immune response in a  
 CC mammal against infection with a virulent Mycobacterium. (I) and (II) are  
 CC also useful as diagnostic reagent for the diagnosis of a virulent  
 CC Mycobacterium infection. The vaccine of the invention induces efficient  
 CC immunological memory, providing long term protection against TB. This  
 CC sequence represents a Microbacterium tuberculosis TB33 antigen described  
 CC in the invention

CC Sequence 297 AA;

Query Match 59.4%; Score 41; DB 3; Length 297;

Best Local Similarity 45.5%; Pred. No. 3.4e+02; Mismatches 2; Indels 0; Gaps 0;

Matches 5; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 2 ISSYDGSWMDY 12

DB 267 VAVYDGSWSEW 277

Search completed: October 5, 2004, 07:59:59

Job time : 12.1296 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2004, 07:29:35 ; Search time 1.25309 Seconds  
(without alignments)  
581.749 Million cell updates/sec

Title: US-09-805-290A-16  
Perfect score: 79  
Sequence: 1 HTTPAGSSNYVGY 14

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	44	55.7	633	1	C2AD_BACTU
2	43	54.4	307	1	KHSE_DEIRA
3	42	53.2	1736	1	CA2B_MOUSE
4	41	51.9	432	1	TIG_SALTY
5	41	51.9	434	1	TIG_YERPE
6	41	51.9	951	1	XRN2_MOUSE
7	41	51.9	976	1	VP41_ARATH
8	40.5	51.3	919	1	SVI_THEMA
9	40	50.6	377	1	AMPC_ECOLI
10	40	50.6	479	1	EFT2_SOYBN
11	40	50.6	746	1	EZH2_MOUSE
12	40	50.6	746	1	EZH2_MOUSE
13	40	50.6	1186	1	CEAA_BACTS
14	39.5	50.0	620	1	TERM_ADDB2
15	39	49.4	223	1	CCGI_MOUSE
16	39	49.4	369	1	RP2_CHUPN
17	39	49.4	519	1	YNE7_HAEIN
18	39	49.4	537	1	YNE7_CAEEL
19	39	49.4	731	1	MCCA_SOYBN
20	39	49.4	743	1	TLE2_HUMAN
21	39	49.4	767	1	TLE2_MOUSE
22	39	49.4	1302	1	RPOB_SPICI
23	38.5	48.7	336	1	MEAY_STRP3
24	38.5	48.7	336	1	MEAY_STRP8
25	38.5	48.7	336	1	MEAY_STRPY
26	38	48.1	402	1	YDZ9_SCHPO
27	38	48.1	506	1	CDYM_HUMAN
28	38	48.1	619	1	Y817_ARCFU
29	38	48.1	641	1	IMD_ARIGO
30	38	48.1	1972	1	P53I_HUMAN
31	37.5	47.5	219	1	P5B9_MOUSE
32	37	46.8	103	1	Y273_METUA
33	37	46.8	125	1	NTF2_YEAST

ALIGNMENTS

Query Match	Similarity	Score	DB	Length	Mismatches	Gaps
3	TPAGSSNYVGY 14	55.7%	DB 1	633	0	0
410	TPRGNISNYVGY 421	66.7%	DB 1	421	3	0

RESULT 1  
C2AD\_BACTU STANDARD; PRT; 633 AA.  
AC Q9RMG3;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Peptide sequence of a new Bacillus thuringiensis cry2-type gene.  
DE CRYIIA(d) (Crystal protein entomocidal protoxin) (71 kDa crystal protein).  
DE CRYIAD OR CRYIAD) OR CRY2.  
OS Bacillus thuringiensis.  
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
OX NCBI\_TaxID=1428;  
RN [1]  
RP STRAIN=BR30;  
RC SEQUENCE FROM N.A.  
RA Choi S.-K., Shin B.-S., Park S.-H.;  
RT "Nucleotide sequence of a new Bacillus thuringiensis cry2-type gene."  
RT Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  
RL FUNCTION: PROMOTES COLLOIDSMOTIC LYSIS BY BINDING TO THE MIDGUT  
CC EPIITHEIAL CELLS OF INSECTS.  
CC DEVELOPMENTAL STAGE: The crystal protein is produced during  
CC sporulation and is accumulated both as an inclusion and as part of  
CC the spore coat.  
CC MISCELLANEOUS: Toxic segment of the protein is located in the N-terminus.  
CC SIMILARITY: Belongs to the delta endotoxin family.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-eb.ch/announce/>  
CC or send an email to [license@isb-eb.ch](mailto:license@isb-eb.ch)).  
CC EMBL: AF200816; AAF09583.1;  
CC InterPro: IPR001178; Endotoxin.  
CC InterPro: IPR005638; Endotoxin.  
CC InterPro: IPR005639; Endotoxin.  
CC InterPro: IPR008979; Gal\_bind\_like.  
CC Pfam: PF03944; endotoxin\_C.1.  
CC Pfam: PF03945; endotoxin\_N.1.  
CC Toxin; Sporulation.  
KW SEQUENCE 633 AA; 70752 MW; 2A58206731B39CB CRC64;  
SQ

```

KHSE DEIRA
ID KHSE DEIRA STANDARD; PRT; 307 AA.
AC Q9R8U5;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Homoserine kinase (EC 2.7.1.39) (HK).
GN THRB OR DR2390.
OS Deinococcus radiodurans.
OC Bacteria; Deinococcus-Thermus; Deinococci; Deinococcales;
OC Deinococcaceae; Deinococcus.
OX NCBI_TaxId=1299;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=R1 / ATCC 13939 / DSM 20539 / NCIB 9279;
RX MEDLINE=20036896; PubMed=10567266;
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zaleski C.,
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RT "Genome sequence of the radioresistant bacterium Deinococcus
RT radiodurans R1."
RL Science 286:1571-1577(1999).
CC -1- CATALYTIC ACTIVITY: ATP + L-homoserine = ADP + O-phospho-L-
CC homoserine.
CC -1- PATHWAY: Threonine biosynthesis from aspartate; fourth step.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC -1- SIMILARITY: Belongs to the GMP kinase family. Homoserine kinase
CC subfamily.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE002069; AAF1935.1; -
DR PIR; G75280; G75280.
DR TIGR; DR2390; -
DR HAMAP; MF_00384; -; 1.
DR InterPro; IPR006204; GMP_Kinase.
DR InterPro; IPR006203; GMP_Kinase.
DR InterPro; IPR000870; Homoser_Kin.
DR Pfam; PF00286; GMP_Kinases; 1.
DR PRINTS; PR00358; HOMSERKINASE.
DR TIGRFAMs; TIGR00191; thrb; 1.
DR PROSITE; PS00627; GMP_KINASES_ATP; 1.
KW Threonine biosynthesis; Transferrase; Kinase; ATP-binding;
KM Complete proteome.
FT NP_BIND 91 101 ATP (POTENTIAL).
SO SEQUENCE 307 AA; 32389 MW; 224E76A7B5E334C CRC64;
Oy 3 TPAGSSNYVY 12
Db 56 TPADESNVYV 65
Query Match 54.4%; Score 43; DB 1; Length 307;
Best Local Similarity 80.0%; Pred. No. 7.2;
Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

DE Collagen alpha 2(XI) chain precursor.
GN COL1A2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 7).
RC STRAIN=129/Sv;
RA Rowen L., Qin S., Madan A., Loretz C., James R., Dors M., Mix L.,
RA Hall J., Lasky S., Hood L.;
RT "Sequence of the mouse major histocompatibility locus class II
RT region."
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 1-1678 FROM N.A. (ISOFORM 7).
RC STRAIN=FVB/N, and 129/Sv; TISSUE=Cartilage;
RX MEDLINE=97135795; PubMed=8981332;
RA Vandenberg P., Voorstee M.W., Ala-Kokko L., Prockop D.J.;
RT "The mouse colla2 gene. Some transcripts from the adjacent rrx-beta
RT gene extend into the colla2 gene."
RL Matrix Biol. 15:359-367(1996).
RN [3]
RP SEQUENCE OF 1-624 FROM N.A. (ISOFORMS 1; 2; 3; 4; 5 AND 6).
RC STRAIN=129/Sv;
RX MEDLINE=95138212; PubMed=7836472;
RA Tsunaki N., Kimura T.;
RT "Differential expression of an acidic domain in the amino-terminal
RT propeptide of mouse pro-alpha2(XI) collagen by complex alternative
RT splicing."
RL J. Biol. Chem. 270:2372-2378(1995).
RN [4]
RP SEQUENCE OF 1-8 FROM N.A.
RC STRAIN=129/Sv; TISSUE=Liver;
RX MEDLINE=96427460; PubMed=8830784;
RA Tsunaki N., Kimura T., Matsui Y., Ochi T.;
RT "Separable cis-regulatory elements that contribute to tissue- and
RT site-specific alpha 2(XI) collagen gene expression in the embryonic
RT mouse cartilage."
RL J. Cell Biol. 134:1573-1582(1996).
RN [5]
RP -1- FUNCTION: May play an important role in fibrillogenesis by
RP controlling lateral growth of collagen II fibrils (by similarity).
RP -1- SUBUNIT: Trimer composed of three different chains: alpha 1(XI),
RP alpha 2(XI), and alpha 3(XI). Alpha 3(XI) is a post-translational
RP modification of alpha 1(XI). Alpha 1(XI) can also be found instead
RP of alpha 3(XI)-1(II) (by similarity).
RN [6]
RP -1- ALTERNATIVE PRODUCTS:
RP Event=Alternative splicing; Named isoforms=7;
RP Comment=Additional isoforms seem to exist;
RP Name=1; Synonyms=E56789;
RP IsoId=Q64739-2; Sequence=Displayed;
RP Name=2; Synonyms=E5689;
RP IsoId=Q64739-3; Sequence=VSP_007346;
RP Name=3; Synonyms=E5789;
RP IsoId=Q64739-4; Sequence=VSP_007345;
RP Name=4; Synonyms=E569;
RP IsoId=Q64739-5; Sequence=VSP_007347;
RP Name=5; Synonyms=E589;
RP IsoId=Q64739-6; Sequence=VSP_007345, VSP_007346;
RP Name=6; Synonyms=E59;
RP IsoId=Q64739-7; Sequence=VSP_007345, VSP_007346, VSP_007347;
RP Name=7;
RP IsoId=Q64739-1; Sequence=VSP_007345, VSP_007347;
RP -1- PTM: Prolines at the third position of the tripeptide repeating
RP unit (G-X-Y) are hydroxylated in some or all of the chains.
RN [7]
RP -1- SIMILARITY: BELONGS TO THE FIBRILLAR CLASS OF COLLAGENS.
RN [8]
RP -1- SIMILARITY: Contains 1 TSP N-terminal (TSPN) domain.
RN [9]
RP This SWISS-PROT entry is copyright. It is produced through a collaboration
RP between the Swiss Institute of Bioinformatics and the EMBL Outstation -
RP the European Bioinformatics Institute. There are no restrictions on its
RP use by non-profit institutions as long as its content is in no way
RP modified and this statement is not removed. Usage by and for commercial

```

entites requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).

CC EMBL, AF100956; AAC69905.1; -.

CC EMBL, U16789; AAA67751.1; -.

DR EMBL, U16790; AAA67752.1; -.

DR EMBL, D38412; BAA18910.1; -.

DR EMBL, D84066; BAA12208.1; -.

DR MGD, MG1:88447; Coll1a2.

DR InterPro, IPR008161; C1g\_helix.

DR InterPro, IPR008160; Collagen.

DR InterPro, IPR008985; Coma\_like\_1ec\_g1.

DR InterPro, IPR001791; Laminin\_G.

DR InterPro, IPR003129; TSPN.

DR Pfam, PF01410; COLFI.1.

DR Pfam, PF01391; Collagen; 18.

DR Pfam, PF02210; TSPN; 1.

DR ProDom, PD000007; C1g\_helix; 1.

DR ProDom, PD002078; Fib\_collagen\_C; 1.

DR SMART, SM00038; COLFI.1.

DR SMART, SM00282; LamG; 1.

DR SMART, SM00210; TSPN; 1.

KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation; Glycoprotein; Collagen; Alternative splicing; Signal.

KW POTENTIAL.

FT CHAIN 1 22

FT PROPE 23 1500

FT DOMAIN 1501 1736

FT DOMAIN 31 214

FT DOMAIN 215 466

FT DOMAIN 487 1500

FT DOMAIN 1501 1736

FT CARBOHYD 1604 1604

FT VARSPIC 267 292

FT VARSPIC 293 313

FT VARSPIC 314 373

FT CONFLICT 536 536

FT CONFLICT 621 621

FT CONFLICT 704 705

FT CONFLICT 797 797

FT CONFLICT 843 845

FT CONFLICT 854 854

FT CONFLICT 876 876

FT CONFLICT 889 889

FT CONFLICT 922 922

FT CONFLICT 1005 1005

FT CONFLICT 1253 1253

FT CONFLICT 1386 1386

FT CONFLICT 1522 1522

FT SEQUENCE 1736 AA; 171535 MW; 18D792D4A387C61 CRC64;

Query Match 53.2%; Score 42; DB 1; Length 1736;

Best Local Similarity 58.3%; Pred. No. 57;

Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 3 TPAGSSNVYGY 14

DB 339 SPAGFYDYTYG 350

RESULT 4

TIG\_SALTY STANDARD; PRT; 432 AA.

AC 08KCA;

DT 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Trigger factor (TF)

GN TIG OR STW0447 OR STY0489 OR T2413.

OS *Salmonella typhimurium*, and

OS *Salmonella typhi*.

CC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

CC Enterobacteriaceae; *Salmonella*.

OK NCBI\_TaxID=602, 601;

RN [1]

RP SEQUENCE FROM N.A.

RC SPECIES=S.typhimurium; STRAIN=LT2 / SGGC1412 / ATCC 700720;

RX MEDLINE=21534948; PubMed=11677609;

RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P., Courtney L., Portwolk S., Ali J., Dante M., Du F., Hou S., Layman D., Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E., Ryan E., Sun H., Flores L., Miller W., Stoneking T., Nhan M., Waterston R., Wilson R.K.;

RT "Complete genome sequence of *Salmonella enterica* serovar Typhimurium LT2.";

RT Nature 413:852-856(2001).

RN [2]

RP SEQUENCE FROM N.A.

RC SPECIES=S.typhi; STRAIN=CT18;

RX MEDLINE=21534947; PubMed=11677608;

RA Parthill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J., Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M., Baker S., Baatman D., Brooke K., Chillingworth T., Connor P., Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J., Felwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jazels K., Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C., Quail M.A., Rutherford K., Simmonds M., Skelton J., Stevens K., Whitehead S., Barrall B.G.;

RT "Complete genome sequence of a multiple drug resistant *Salmonella enterica* serovar Typhi CT18.";

RT Nature 413:848-852(2001).

RN [3]

RP SEQUENCE FROM N.A.

RC SPECIES=S.typhi; STRAIN=Ty2 / ATCC 700931;

RX MEDLINE=22531367; PubMed=12644504;

RA Deng W., Iou S.R., Plunkett G. III, Mayhew G.F., Rose D.J., Burdand V., Kodymani V., Schwartz D.C., Blatner F.R.;

RT "Comparative genomics of *Salmonella enterica* serovar Typhi strains Ty2 and CT18.";

RT J. Bacteriol. 185:2330-2337(2003).

RT -I- FUNCTION: INVOLVED in protein export. Acts as a chaperone by maintaining the newly synthesized protein in an open conformation (By similarity).

CC -I- SIMILARITY: Belongs to the FKBP-type piase family. Tlg subfamily.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

CC EMBL, AE008716; AAL19402.1; -.

DR EMBL, AL627266; CAD08906.1; -.

DR EMBL, AE016842; AAO70003.1; -.

DR StvGene; SG77777; tlg.

DR HAMAP, MF\_00303; -; 1.

DR InterPro, IPR001179; FKBP\_piase.

DR InterPro, IPR005215; Tlg\_fac.

DR InterPro, IPR008880; Tlg9er\_C.

DR InterPro, IPR008881; Tlg9er\_N.

DR Pfam, PF00254; FKBP; 1.

DR Pfam, PF05698; Tlg9er\_C; 1.

DR Pfam, PF05697; Tlg9er\_N; 1.

DR TIGFAME, TIGR00115; Fig; 1.

DR PROSITE, PS00453; FKBP\_PIASE\_1; FALSE\_NRG.

DR PROSITE, PS00454; FKBP\_PIASE\_2; FALSE\_NRG.

DR PROSITE, PS50059; FKBP\_PIASE\_3; 1.

KW Cell division; Chaperone; Isomerase; Rotamase; Complete proteome.

FT DOMAIN 161 246 PPIASE, FKBP-TYPE  
SQ SEQUENCE 432 AA, 48066 MW, B6636BF9B8E7CA7C CRC64;  
Query Match 51.9%; Score 41; DB 1; Length 432;  
Best Local Similarity 66.7%; Pred. No. 22;  
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
QY 2 ITPAGSNVYVG 13  
DB 84 INPACAPNTVPG 95

RESULT 5  
TIG\_YERPE  
ID TIG\_YERPE STANDARD; PRT; 434 AA.  
AC 082C64;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DE 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Trigger factor (TF)  
GN TIG OR YPO3158 OR Y1026.  
OS Yersinia pestis.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Yersinia.  
OX NCBI\_TaxID=632;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CO-92 / Biovar Orientalis;  
RX MEDLINE=21470413; PubMed=1156360;  
RA Patchell J, Wren B.W., Thomson N.R., Tibball R.W., Holden M.T.G.,  
RA Pretlie M.B., Sebainia M., James K.D., Churcher C., Mungall K.L.,  
RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdano-Tarrega A.M.,  
RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,  
RA Feltham T., Hamlin N., Holtroft S., Jagals K., Karlyshev A.V.,  
RA Leather S., Moulis S., Oyston P.C.F., Quail M.A., Rutherford K.,  
RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrall B.G.,  
RT "Genome sequence of Yersinia pestis, the causative agent of plague."  
RL Nature 413:523-527(2001).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=KIMS / Biovar Mediaevalis;  
RX MEDLINE=22137863; PubMed=12142430;  
RA Deng W., Burland V., Plunkett G. III, Boutin A., Mayhew G.F., Liss P.,  
RA Pena N.T., Rose D.J., Mau B., Zhou S., Schwartz D.C.,  
RA Fetherston J.D., Lindler L.E., Brubaker R.R., Plano G.V.,  
RA Straley S.C., McDonough K.A., Nilles M.L., Watson J.S., Blattner F.R.,  
RA Perry R.D.;  
RT "Genome sequence of Yersinia pestis KIM."  
RL J. Bacteriol. 184:4601-4611(2002).  
CC -!- FUNCTION: Involved in protein export. Acts as a chaperone by  
maintaining the newly synthesized protein in an open conformation  
(By similarity).  
CC -!- SIMILARITY: Belongs to the FKBP-type PPIase family. Tlg subfamily.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
DR EMBL/ AJ14155; CAC92393.1; -  
DR EMBL/ AE013706; AAM84607.1; -  
DR PIR/ AP0383; AP0383.  
DR HAMAP/ MF\_00303; -; 1.  
DR InterPro/ IPR001179; FKBP\_PPIase.  
DR InterPro/ IPR005215; Tlg\_fac.  
DR InterPro/ IPR008880; Trigger\_C.  
DR InterPro/ IPR008881; Trigger\_N.  
DR Pfam/ PF00254; FKBP.1.  
DR Pfam/ PF05698; Trigger\_C.1.  
DR Pfam/ PF05699; Trigger\_N.1.

DR TIGRfam: TIGR00115; tlg.1  
DR PROSITE/ PS00453; FKBP\_PPIASE\_1; FALSE\_NEG.  
DR PROSITE/ PS00454; FKBP\_PPIASE\_2; FALSE\_NEG.  
DR PROSITE/ PS00509; FKBP\_PPIASE\_3; 1.  
KM Cell division: Chaperone, isomerase, Rotamase; Complete proteome.  
FT DOMAIN 161 246 PPIASE, FKBP-TYPE  
SQ SEQUENCE 434 AA, 48240 MW, D44CB4FB9A5CAB CRC64;  
Query Match 51.9%; Score 41; DB 1; Length 434;  
Best Local Similarity 66.7%; Pred. No. 22;  
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
QY 2 ITPAGSNVYVG 13  
DB 84 INPACAPNTVPG 95

RESULT 6  
XRN2\_MOUSE  
ID XRN2\_MOUSE STANDARD; PRT; 951 AA.  
AC 09DFL; 061489; 093X87;  
DT 10-OCT-2003 (Rel. 42, Created)  
DT 10-OCT-2003 (Rel. 42, Last sequence update)  
DE 15-MAR-2004 (Rel. 43, Last annotation update)  
DE 5'-3' exoribonuclease 2 (EC 3.1.11.-) (Dhml protein).  
GN XRN2 OR DHML.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORM 2).  
RC TISSUE=Testis;  
RX MEDLINE=95192042; PubMed=7868930;  
RA Shobnike T., Sugano S., Yamashita T., Ikeda H.;  
RT "Characterization of cDNA encoding mouse homolog of fission yeast  
dhp1 gene: structural and functional conservation."  
RL Nucleic Acids Res. 23:357-361(1995).  
RN [2]  
RP SEQUENCE FROM N.A. (ISOFORM 1).  
RC STRAIN=C57BL/6J; TISSUE=Embryo. Eye, Forelimb, and Lung;  
RX MEDLINE=22554683; PubMed=12466851;  
RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,  
RA Nishida I., Oshio N., Saito R., Suzuki H., Yamana H., Kiyosawa H.,  
RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojohori T.,  
RA Balarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,  
RA Schriml L.M., Kanapin A., Matsuda H., Batilov S., Beisel K.W.,  
RA Blake J.A., Brad D., Brusic V., Chochia C., Corbani L.E., Cousins S.,  
RA Dalla E., Dragani T.A., Fletcher C.P., Forrest A., Frazer K.S.,  
RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,  
RA Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,  
RA Kanai A., Kawai H., Kawasawa Y., Kedzierski R.M., King B.L.,  
RA Kongaya A., Kurochkin I.V., Lee Y., Lemhard B., Lyons P.A.,  
RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,  
RA Nagashima T., Nunata K., Okido T., Pavan W.J., Petrea G.,  
RA Petrovsky N., Pillai R., Pontus J.V., Qi D., Ramachandran S.,  
RA Ravasi T., Reed J.C., Reed D.O., Reid J., Ring B.Z., Ringwald M.,  
RA Sandelin A., Schneider C., Sempile C.A., Setou M., Shimada K.,  
RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,  
RA Verardo R., Wagner L., Walestedt C., Wang Y., Watanabe Y., Wells C.,  
RA Wilting L.G., Wyshaw-Boris A., Yangisawa M., Yang I., Yang L.,  
RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carlinici P., Hayatsu N.,  
RA Hirozane-Kishikawa T., Kono H., Nakamura M., Sakazume N., Sato K.,  
RA Shiraki T., Waki K., Kawai J., Aizawa K., Arikawa T., Fukuda S.,  
RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,  
RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,  
RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,  
RA Birney E., Hayashizaki Y.;  
RT "Analysis of the mouse transcriptome based on functional annotation of  
60,770 full-length cDNAs."  
RL Nature 420:563-573(2002).  
RN [3]  
RP SEQUENCE FROM N.A. (ISOFORM 1).



```

CC -----
DR EMBL, AC011438; AAF18239.1; ALT_SEQ.
DR EMBL, U86653; AAB60858.1; -.
DR InterPro; IPR000547; Clathrin_repeat.
DR InterPro; IPR001680; WD40.
DR pfam; PF00637; Clathrin_1.
DR pfam; PF00400; WD40; 1.
DR SMART; SM00299; CLH; 1.
DR SMART; SM00184; RING; 1.
DR SMART; SM00320; WD40; 1.
DR PROSITE; PS00682; WD_REPEATS_1; FALSE_NEG.
DR PROSITE; PS00682; WD_REPEATS_2; 1.
DR PROSITE; PS00294; WD_REPEATS_REGION; FALSE_NEG.
KW Repeat; WD repeat.
FT REPEAT 82 121 WD 2.
FT REPEAT 175 211 WD 2.
FT DOMAIN 15 33 POLY-GLU.
FT DOMAIN 941 950 POLY-GLU.
FT DOMAIN 970 975 POLY-ALA.
FT CONFLICT 644 644 R -> G (IN REF. 2).
FT CONFLICT 944 944 E -> EE (IN REF. 2).
SQ SEQUENCE 976 AA; 109854 MM; C9FEB68040869D12 CRC64;

Query Match
Best Local Similarity 51.9%; Score 41; DB 1; Length 976;
Matches 8; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 1 HITPAGSSNYVGY 14
DB 922 HKTGSSGYEYSY 935

RESULT 8
SYI THEME STANDARD; PRT; 919 AA.
AC P46213;
DT 01-NOV-1995 (Rel. 32, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Isolucyl-tRNA synthetase (EC 6.1.1.5) (Isolucine--tRNA ligase)
DE (Ilers).
GN ILRS OR TM1361.
OS Thermotoga maritima.
OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
CX NCBI_TaxID=2336;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MSB8 / DSM 3109 / ATCC 43589;
RX MEDLINE=99287316; PubMed=10360571;
RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RT "Evidence for lateral gene transfer between Archaea and Bacteria from
RT genome sequence of Thermotoga maritima.";
RT Nature 399:323-329 (1999).
RN [2]
RP SEQUENCE OF 95-599 FROM N.A.
RC MEDLINE=95223956; PubMed=7708661;
RA Brown J.R., Doclittle W.F.;
RT "Root of the universal tree of life based on ancient aminoacyl-tRNA
RT synthetase gene duplications.";
RT Proc. Natl. Acad. Sci. U.S.A. 92:2441-2445 (1995).
CC -1- CATALYTIC ACTIVITY: ATP + L-isoleucine + tRNA(Ile) = AMP +
CC diphosphate + L-isoleucyl-tRNA(Ile).
CC -1- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC -1- SUBUNIT: Monomer (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC -----

```

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, AB001790; AAD36431.1; -.
DR EMBL, L37104; AAC14448.1; -.
DR PIR; B72263; B72263.
DR HSSP; PA1972; 1FFV.
DR TIGR; TM1361; -.
DR InterPro; IPR002300; tRNA-synt_1a.
DR InterPro; IPR001412; tRNA-synt_1.
DR InterPro; IPR002301; tRNA-synt_1le.
DR InterPro; IPR009008; Valrs_1lers_edit.
DR pfam; PF00133; tRNA-synt_1; 1.
DR PRINTS; PRO0984; TRNASYNTHILE.
DR TIGRPFAM; TTGR00392; 11es; 1.
DR PROSITE; PS00178; AA_tRNA_LIGASE_1; 1.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW Metal-binding; Zinc; Complete proteome.
FT SITE 57 67 "HIGH" REGION.
FT SITE 594 598 "KMSKS" REGION.
FT BINDING 597 597 ATP (By similarity).
SQ SEQUENCE 919 AA; 107155 MM; 40E4D0876010C385 CRC64;

Query Match
Best Local Similarity 51.3%; Score 40.5; DB 1; Length 919;
Matches 7; Conservative 3; Mismatches 4; Indels 1; Gaps 1;

OY 1 HITPA-GSSNYVGY 14
DB 320 HAPGHEEDYHYGH 334

RESULT 9
AMPC ECOLI
ID AMPC ECOLI STANDARD; PRT; 377 AA.
AC P00811;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Beta-lactamase precursor (EC 3.5.2.6) (Cephalosporinase).
GN AmpC OR AMPA OR B4150.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
CX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 20-31.
RC STRAIN=K12;
RX MEDLINE=82060161; PubMed=6795623;
RA Jaurin B., Grundstroem T.;
RT "AmpC cephalosporinase of Escherichia coli K-12 has a different
RT evolutionary origin from that of beta-lactamases of the penicillinase
RT type.";
RT Proc. Natl. Acad. Sci. U.S.A. 78:4897-4901 (1981).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655.
RX MEDLINE=9534362; PubMed=7610040.
RA Blatner F.R.;
RA Blatner V.D., Plunkett G. III, Sofia H.J., Daniels D.L.;
RT "Analysis of the Escherichia coli genome VI: DNA sequence of the
RT region from 92.8 through 100 minutes.";
RT Nucleic Acids Res. 23:2105-2119 (1995).
RN [3]
RP SEQUENCE OF 1-29 FROM N.A.
RC MEDLINE=82174546; PubMed=7041115;
RA Grundstroem T., Jaurin B.;
RT "Overlap between ampC and ftd operons on the Escherichia coli

```

RT Chromosome,";  
RL Proc. Natl. Acad. Sci. U.S.A. 79:1111-1115(1982).  
RN [4]  
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).  
RX MEDLINE=99036630; PubMed=9819201;  
RA Usher K.C., Blaszcak L.C., Weston G.S., Shoichet B.K.,  
RA Remington S.U.,  
RT "Three-dimensional structure of AmpC beta-lactamase from Escherichia  
RT coli bound to a transition-state analogue: possible implications for  
RT the oxyanion hypothesis and for inhibitor design.";  
RL Biochemistry 37:16082-16092(1998).  
[5]  
RP X-RAY CRYSTALLOGRAPHY (2.25 ANGSTROMS).  
RX MEDLINE=20060984; PubMed=10595535;  
RA Powers R.A., Blazquez U., Weston G.S., Morosini M.I., Baquero F.,  
RA Shoichet B.K.,  
RT "The complexed structure and antimicrobial activity of a non-beta-  
RT lactam inhibitor of AmpC beta-lactamase.";  
RL Protein Sci. 8:2330-2337(1999).  
CC -I- FUNCTION: THIS PROTEIN IS A SERINE BETA-LACTAMASE WITH A SUBSTRATE  
CC SPECIFICITY FOR CEPHALOSPORINS.  
CC -I- CATALYTIC ACTIVITY: A beta-lactam + H(2)O = a substituted beta-  
CC amino acid.  
CC -I- SUBUNIT: Monomer.  
CC -I- SUBCELLULAR LOCATION: Periplasmic.  
CC -I- SIMILARITY: Belongs to the class-C beta-lactamase family.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
-----  
DR EMBL; U01611; AAA23441.1; -  
DR EMBL; U14003; AAA97049.1; -  
DR EMBL; AE000487; AAC77110.1; -  
DR EMBL; V00277; CAA23537.1; -  
DR PIR; A01007; OKEC.  
DR PDB; 2BIS; 12-AUG-98.  
DR PDB; 3BIS; 12-AUG-98.  
DR PDB; 1C3B; 24-NOV-99.  
DR PDB; 1FCM; 18-DEC-00.  
DR PDB; 1FCN; 04-DEC-00.  
DR PDB; 1FSY; 14-MAR-01.  
DR PDB; 1GAY; 25-JUL-01.  
DR PDB; 1IEI; 15-AUG-01.  
DR PDB; 1IEM; 15-AUG-01.  
DR PDB; 1KDS; 17-JUL-02.  
DR PDB; 1KEJ; 17-JUL-02.  
DR PDB; 1KE3; 17-JUL-02.  
DR PDB; 1KE4; 17-JUL-02.  
DR PDB; 1LIS; 27-NOV-02.  
DR PDB; 1LIS; 02-OCT-02.  
DR PDB; 1LIS; 02-OCT-02.  
DR PDB; 1LIS; 02-OCT-02.  
DR PDB; 1MXO; 04-MAR-03.  
DR PDB; 1MX8; 04-MAR-03.  
DR MEROPS; S12.0NM; -  
DR SWISS-2DPAGE; P00811; COLI.  
DR ECO2DBASE; I035.7; 6TH EDITION.  
DR EcoGene; EGI0040; ampC.  
DR InterPro; IPR001586; Beta\_lactam\_C\_AS.  
DR InterPro; IPR001466; Beta\_lactamase.  
DR InterPro; IPR000437; Prok\_lipoprot\_S.  
DR Pfam; PF00144; beta\_lactamase\_1.  
DR PROSITE; PS00336; BETA\_LACTAMASE\_C; 1.  
KW Hydrolyse; Antibiotic resistance; Periplasmic; Signal; 3D-structure;  
KW Complete proteome.

FT	SIGNAL	1	19	
FT	CHAIN	20	377	
FT	ACT_SITE	80	80	
FT	BINDING	331	333	
FT	HELIX	22	39	
FT	TURN	40	40	
FT	STRAND	43	50	
FT	TURN	51	52	
FT	STRAND	53	63	
FT	TURN	64	67	
FT	STRAND	66	69	
FT	TURN	72	73	
FT	STRAND	75	77	
FT	HELIX	79	81	
FT	HELIX	82	95	
FT	TURN	96	97	
FT	TURN	101	102	
FT	STRAND	104	104	
FT	HELIX	105	107	
FT	TURN	108	108	
FT	TURN	110	111	
FT	HELIX	115	117	
FT	TURN	118	119	
FT	STRAND	121	121	
FT	HELIX	122	127	
FT	TURN	128	128	
FT	TURN	139	140	
FT	HELIX	144	153	
FT	TURN	160	161	
FT	STRAND	163	165	
FT	HELIX	168	178	
FT	TURN	179	183	
FT	HELIX	186	193	
FT	TURN	194	194	
FT	HELIX	195	198	
FT	TURN	199	199	
FT	TURN	201	202	
FT	STRAND	204	204	
FT	HELIX	209	214	
FT	STRAND	215	215	
FT	STRAND	218	220	
FT	TURN	221	222	
FT	STRAND	223	225	
FT	TURN	231	232	
FT	HELIX	233	237	
FT	STRAND	240	241	
FT	HELIX	243	254	
FT	HELIX	256	258	
FT	HELIX	262	271	
FT	TURN	272	272	
FT	STRAND	274	278	
FT	TURN	279	280	
FT	STRAND	281	283	
FT	STRAND	288	291	
FT	HELIX	296	302	
FT	TURN	303	303	
FT	HELIX	305	308	
FT	TURN	309	309	
FT	STRAND	311	313	
FT	STRAND	315	321	
FT	STRAND	327	335	
FT	TURN	336	337	
FT	STRAND	338	345	
FT	HELIX	346	348	
FT	TURN	349	349	
FT	STRAND	350	356	
FT	STRAND	358	358	
FT	HELIX	362	375	
SQ	SEQUENCE	377 AA;	41555 MW;	3CGFBAFEAEF96C9F CAC64;

Query Match 50.6%; Score 40; DB 1; length 377;  
Best Local Similarity 42.9%; Pred. No. 28;  
Matches 6; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 1 HITPAGSSNYVY 14  
DB 206 NVPFAEKXVAMGY 219

RESULT 10  
EFT2\_SOYBN STANDARD; PRT; 479 AA.  
ID EFT2\_SOYBN  
AC P46280; 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE Elongation factor Tu, chloroplast precursor (EF-Tu).  
GN TUBR1.  
OS Glycine max (Soybean).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.  
OX NCBI\_TaxID=3847;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Ceresia;  
RA Maurer P., Murone M., Stutz E.;  
RT "The tuF gene family of soybean: structure and differential  
expression.";  
RL Plant Sci. 117:83-92(1996).  
CC -!- FUNCTION: This protein promotes the GTP-dependent binding of  
aminoacyl-tRNA to the A-site of ribosomes during protein  
biosynthesis.  
CC -!- SUBCELLULAR LOCATION: Chloroplast.  
CC -!- SIMILARITY: Belongs to the GTP-binding elongation factor family.  
EF-Tu/EF-1A subfamily.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; X89058; CAAG1444.1; -;  
DR PIR; S60659; S60659.  
DR HSBP; P02990; IEFU.  
DR InterPro; IPR004541; EF-Tu.  
DR InterPro; IPR000795; EF\_GTPbind.  
DR InterPro; IPR004160; EFTU\_Cterm.  
DR InterPro; IPR006161; EFTU\_D2.  
DR InterPro; IPR009001; Elong\_init\_C.  
DR InterPro; IPR005225; Small\_GTP.  
DR InterPro; IPR009000; Translat\_factor.  
DR Pfam; PF00009; GTP\_EFTU; 1.  
DR Pfam; PF03144; GTP\_EFTU\_D2; 1.  
DR Pfam; PF03143; GTP\_EFTU\_D3; 1.  
DR PRINTS; PR00315; ELONGATINCT.  
DR TIGRPFAMs; TIGR00485; EF-Tu; 1.  
DR TIGRPFAMs; TIGR00231; small\_GTP; 1.  
DR XMAPSIITE; PS00301; EFACOR\_GTP; 1.  
KM Elongation factor; protein biosynthesis; GTP-binding;  
KM Transist peptide; Chloroplast; Multigene family.  
FT TRANSIT 1 70  
FT CHAIN 1 479  
FT NP\_BIND 89 96 ELONGATION FACTOR TU.  
FT NP\_BIND 89 96 GTP (BY SIMILARITY).  
FT NP\_BIND 151 155 GTP (BY SIMILARITY).  
FT NP\_BIND 206 209 GTP (BY SIMILARITY).  
SQ SEQUENCE 479 AA; 52509 MW; CD65F0R262BBD77B CRC64;

Query Match 50.6%; Score 40; DB 1; Length 479;  
Best Local Similarity 50.0%; Pred. No. 35;  
Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 HITPAGSSNYVY 12

DB 20 HSPSPSSNYLF 31

RESULT 11  
EZH2\_HUMAN STANDARD; PRT; 746 AA.  
ID EZH2\_HUMAN  
AC Q15755; Q15755; Q92857;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Enhancer of zeste homolog 2 (ENX-1).  
GN EZH2.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Brain;  
RX MEDLINE=97124843; PubMed=8954776;  
RA Chen H., Rossier C., Antonarakis S.E.;  
RT "Cloning of a human homolog of the Drosophila enhancer of zeste gene  
EZH2 that maps to chromosome 21q22.2.";  
RL Genomics 38:30-37(1996).  
-----  
CC -!- FUNCTION: May be involved in the regulation of gene transcription  
and chromatin structure.  
CC -!- SUBUNIT: BINDS ATRX VIA THE SET DOMAIN (PROBABLE).  
CC -!- SUBCELLULAR LOCATION: Nuclear (probable).  
CC -!- TISSUE SPECIFICITY: Expressed in many tissues.  
CC -!- SIMILARITY: Contains 1 SET domain.  
CC -!- SIMILARITY: Belongs to the EZ family.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; X95653; CAAG4955.1; -;  
DR EMBL; U61145; AAC50591.1; -;  
DR PIR; U52965; AAC50591.1; -;  
DR PIR; G02838; G02838.  
DR TRANSPAC; T04886; -;  
DR Genew; HGNC:3527; EZH2.  
DR MIM; 601573; -;  
GO; GO:0003677; F-DNA binding; TAS.



DR	GO: 0006325; P: establishment and/or maintenance of chromatin; TAS
DR	GO: 0006325; P: establishment and/or maintenance of chromatin; TAS
DR	InterPro: IPR001005; Myb_DNA_binding.
DR	InterPro: IPR001214; SET.
DR	PFam: PF00856; SET; 1.
DR	SMART: SM00717; SANT; 2.
DR	SMART: SM00317; SET; 1.
DR	PROSITE: PS50280; SET; 1.
KW	Transcription regulation; Nuclear protein; DNA-binding; NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT	DOMAIN 490 485
FT	DOMAIN 523 605
FT	DOMAIN 611 731
FT	DOMAIN 724 724
FT	CONFLICT 724 724 F -> L (IN REF. 1).
FT	CONFLICT 724 724 F -> V (IN REF. 1).
SO	SEQUENCE 746 AA; 85363 MW; 1B5029EB9D509B5 CRC64;

Query Match	50.6%;	Score 40;	DB 1;	Length 746;
Best Local Similarity	77.8%;	Pred. No. 54;		
Matches 7; Conservative		1; Mismatches	1; Indels	0; Gaps 0

Qy 6 GSSNIVYGY 14  
|||:|  
Db 512 GSSNIVYNY 520

RESULT 12		
EZH2 MOUSE		
ID	EZH2 MOUSE	STANDARD;
00110		PRT; 746 AA
000000		

DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Enhancer of zeste homolog 2 (ENX-1).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus  
OX NCBI\_TaxID=10090;

RC TISSUE=Placenta;  
RX MEDLINE=97014262; PubMed=8861097;  
RA Robert O., Sures I., Ciossek T., Fuchs M., Ullrich A.,

RP SEQUENCE OF 134-497 FROM N. J.

RA Laible G., Haynes A.R., Lebersorger A., O'Carroll D., Mattei M.-G.,  
RA Denny P., Brown S.D., Jenuwein T.;  
RT "The murine polycomb-group genes *ezh1* and *ezh2* map close to *hox* gene  
RA clusters on chromosomes 12 and 17, respectively." *Proc Natl Acad Sci USA* 95:10513-10518 (1998)

CC -|- SURF1: BINDS ATRX VIA THE SET DOMAIN (BY SIMILARITY)  
CC and chromatin structure.  
CC -|- FUNCTION: May be involved in the regulation of gene transcription  
CC Mamm. genome 10:311-314 (1999).  
KL

```
CC      -I-ALTERNATIVE PRODUCTS;
CC      Event=Alternative splicing; Named isoforms=2
CC      Name=ENX.1A;
CC      TSold=661188-1; Sequence=1 saved;
```

CC Isoid=Q61188-2; Sequence=VSP\_001501;  
CC  
CC -1- TISSUE SPECIFICITY: WIDELY EXPRESSED IN EARLY EMBRYOS. IN LATER  
CC EMBRYOGENESIS, EXPRESSION RESTRICTED TO CENTRAL AND PERIPHERAL  
CC NEBULOSA SYSTEM LAYER AND THINNING IN ADULT VITREOUS MEMBRANE

CC AND VERY LOW LEVELS IN BRAIN AND LIVER. NO EXPRESSION IN HEART,  
CC THYROID GLAND, LUNG AND KIDNEY.  
CC -1- DEVELOPED STAGE: EXPRESSED IN BOTH ADULT AND EMBRYO WITH  
CC HIGHEST LEVELS IN EARLY EMBRYONIC STAGE.

HIGHEST LEVELS IN EARLY EMBRYOGENESIS

CC -1- SIMILARITY: Contains 1 SET domain.

CC -1~ SIMILARITY: Belongs to the EZ family

CC This SWISS-PROF entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce>;  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; U52951; AAC52655.1; -.  
DR EMBL; AF104359; AAD54020.1; -.  
DR MCB; MG1:107940; Ezr2.  
DR GO; GO:0005634; C:nucleus; IDA.

DR InterPro; IPR001005; Myb\_DNA\_binding.  
DR InterPro; IPR001214; SET.  
DR Pfam; PF00856; SET; 1.  
DR SMART; SM00717; SANT; 2.

DR	PROSITE; PS50280; SET; 1.	
KM	Transcription regulation; Nuclear protein; DNA-binding;	
KT	Alternative splicing.	
ET	DOMAIN 490 495	NUCLEAR LOCALIZATION SIGNAL (POTENTIAL)

FT	DOSNAME	723	932
FT	DOMAIN	611	731
FT	VARSPLIC	511	553
FT			

```

CIB NAME:
SET.
DSSSHVNYQPCDHROPCOSSCPVIAQNFCXCFCCCSS
BC -> G (in isoform ENX-1B).
/FTid=VSP 001501.

```

Query Match	50.6%;	Score 40;	DB 1;	Length
Best local similarity	77.8%;	Pred. No. 54;		
CONTACT	135	101	MISSING	1633ED24E CRC64
SEQUENCE	746 AA;	85336 MW;	0435C021933ED24E	CRC64

Matches	Conservative	Mismatches	Indels	Gaps
QY	6 GSSNRYIGY 14			
db	512 GSSNRYTNY 520			

RESULT 13	
CEAA_BACTS	
ID_CEAA_BACTS	STANDARD
	DATE

```
AC Q45710;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Pectinase from Aspergillus niger ATCC 26455
```

DE CryXIVa(a) (Crystalline entomocidal protoxin) (132 kDa crystal protein).  
GN CryI4Aa OR CryXIVa(A).  
OS *Bacillus thuringiensis* (subsp. *sotto*)...

```

OX  NCBI_TaxID=29340;
RN  [1]
RP  SEQUENCE FROM N.A.
EC  STRAIN=NRRL B-18679 / PS800J71.

```

RT "Novel *Saccharum thuringiensis* toxins active against corn rootworm  
RT larvae";  
RI Patent number WO9416079, 21-JUL-1994.  
CC -- FUNCTION: PROMOTES COLLOIDSMOTIC LYSIS BY BINDING TO THE MIDGUT

CC -1- DEVELOPMENTAL STAGE: The crystal protein is produced during  
CC sporulation and is accumulated both as an inclusion and as part of  
CC the spore coat.  
CC  
CC -1- MISCELLANEOUS: Toxic segment of the protein is located in the N-

CC terminus.  
CC  
CC -1- SIMILARITY: Belongs to the delta endotoxin family.  
CC  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U13955; AAA2516.1; -
DR PIR: T18210; T18210.
DR InterPro: IPR001178; Endotoxin.
DR InterPro: IPR005638; endotoxin.C.
DR InterPro: IPR005639; endotoxin.N.
DR InterPro: IPR008979; Gal_bind_like.
DR Pfam: PF00555; endotoxin_1.
DR Pfam: PF03944; endotoxin_C.1.
DR Pfam: PF03945; endotoxin_N.1.
DR Toxin; Sporulation.
DR KX SEQUENCE 1186 AA; 131694 MW; EF4BBIHAEZCB9487 CRC64;

Query Match 50.6%; Score 40; DB 1; Length 1186;
Best Local Similarity 58.3%; Pred. No. 65;
Matches 7; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 3 TPAGSSNVYGY 14
Db 383 TPAGSGVAYPY 394

RESULT 14
TERM ADER2 STANDARD; PRT; 620 AA.
ID TERMADEB2
AC 055438;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE DNA terminal protein (Bellert protein) (PTP protein).
GN PTP.
OS Bovine adenovirus type 2 (Mastadenovirus bos2).
OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.
OX NCBI_TaxID=114429;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=19;
RX MEDLINE=98275674; PubMed=9612727;
RA Okic D., Jegub A., Bautista D., Haj-Ahmad Y.;
RT "Sequence analysis of the terminal protein precursor coding regions
RT from bovine adenovirus serotypes 2 and 3.";
RL Intervirology 40:253-262(1997).
CC -1- FUNCTION: This protein is covalently attached to the terminl of
CC replicating DNA in vivo and nascent DNA synthesized in vitro and
CC may play some role in DNA replication.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF252854; AAB88488.1; -
DR InterPro: IPR003391; Adeno terminal.
DR Pfam: PF02459; Adeno terminal.1.
DR KM DNA replication; Covalent protein-DNA linkage.
FT BINDING 545 545 COVALENT LINKAGE OF VIRAL DNA (BY
FT SIMILARITY).
SQ SEQUENCE 620 AA; 71440 MW; 08AB33B5953F3124 CRC64;

Query Match 50.0%; Score 39.5; DB 1; Length 620;
Best Local Similarity 56.2%; Pred. No. 55;
Matches 9; Conservative 1; Mismatches 3; Indels 3; Gaps 1;

QY 2 ITPAG---SSNVYGY 14

```

```

Db 52 ITPAGGWMKSYGY 67

RESULT 15
CCGI_MOUSE
ID CCGI_MOUSE STANDARD; PRT; 223 AA.
AC 070578;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Voltage-dependent calcium channel gamma-1 subunit (Dihydropyridine-
DE sensitive L-type, skeletal muscle calcium channel gamma subunit).
GN CACNG1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129/SV; TISSUE=Skeletal muscle;
RX MEDLINE=98163244; PubMed=9504716;
RA Wiesenbach U., Bosse-Denecke E., Freise D., Ludwig A., Murakami M.,
RA Hofmann F., Flockert V.;
RT "The structure of the murine calcium channel gamma-subunit gene and
RT protein.";
RL Biol. Chem. 379:45-50(1998).
CC -1- FUNCTION: This protein is a subunit of the dihydropyridine (DHP)
CC sensitive calcium channel. Plays a role in excitation-contraction
CC coupling. The skeletal muscle DHP-sensitive Ca(2+) channel may
CC function only as a multiple subunit complex.
CC -1- SUBUNIT: The L-type calcium channel is composed of five subunits:
CC alpha-1, alpha-2/delta, beta and gamma.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- SIMILARITY: Belongs to the PMP-22 / EMP / MP20 family. CACNG
CC subfamily.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AJ006306; CA06966.1; -
DR MGD: MGI:1206582; Cacng1.
DR InterPro: IPR004031; PMP22_Claudin.
DR InterPro: IPR008368; VDCCgamma.
DR Pfam: PF00822; PMP22_Claudin; 1.
DR PRINTS: PR01792; VDCCGAMMA.
DR PRINTS: PR01601; VDCCGAMMA1.
RX Ionic channel; Transmembrane; Ion transport; Voltage-gated channel;
RX Calcium channel; Glycoprotein; Phosphorylation.
FT TRANSMEM 11 29 POTENTIAL.
FT TRANSMEM 106 130 POTENTIAL.
FT TRANSMEM 141 156 POTENTIAL.
FT TRANSMEM 181 205 POTENTIAL.
FT CARBOHYD 43 43 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 80 80 N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 223 AA; 25120 MW; 3712F856CAFAE5F9 CRC64;

Query Match 49.4%; Score 39; DB 1; Length 223;
Best Local Similarity 50.0%; Pred. No. 25;
Matches 6; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 ITPAGSSNVY 12
Db 72 HTVPSGKNCYSY 83

Search completed: October 5, 2004, 08:01:51.

```

Tue Oct 5 09:51:04 2004

Job time : 3.25309 secs

us-09-805-290a-16.rsp

Page 11



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2004, 07:34:11 ; Search time 7.41049 Seconds  
(without alignments)  
596.081 Million cell updates/sec

Title: US-09-805-290A-16  
Perfect score: 79  
Sequence: 1 HTTPAGSSNYVYGY 14

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

SPTREMBL 25: \*  
1: sp\_archaea: \*  
2: sp\_bacteria: \*  
3: sp\_fungi: \*  
4: sp\_human: \*  
5: sp\_invertebrate: \*  
6: sp\_mammal: \*  
7: sp\_mhc: \*  
8: sp\_organelle: \*  
9: sp\_phage: \*  
10: sp\_plant: \*  
11: sp\_rodent: \*  
12: sp\_virus: \*  
13: sp\_vertebrate: \*  
14: sp\_unclassified: \*  
15: sp\_virus: \*  
16: sp\_bacteriophage: \*  
17: sp\_archaeap: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	46	58.2	117	5	Q94257
2	45	57.0	307	16	Q89YMS
3	45	57.0	1960	10	Q856N2
4	44	55.7	838	12	Q9QTC3
5	43	54.4	150	9	Q859Q1
6	43	54.4	208	2	Q33707
7	43	54.4	402	10	Q9ZUA7
8	43	54.4	337	16	Q8FWT3
9	42	53.2	337	16	Q82410
10	42	53.2	400	16	Q7URZ3
11	42	53.2	505	2	Q8RM70
12	42	53.2	661	5	Q860R6
13	42	53.2	796	5	Q9V7U9
14	41	51.9	156	16	Q99YX7
15	41	51.9	306	17	Q8ZUN9
16	41	51.9	371	17	Q8TMU9

17	41	51.9	503	10	Q7X819	Q7X819 oryza sativ
18	41	51.9	787	16	Q8A3U2	Q8A3U2 bacteroides
19	41	51.9	4671	5	Q9VH97	Q9VH97 diatrophila
20	40.5	51.3	304	16	Q8DXA8	Q8DXA8 synchococc
21	40.5	51.3	653	3	Q96M49	Q96M49 candida alb
22	40	50.6	57	4	Q9BSF7	Q9BSF7 homo sapien
23	40	50.6	107	10	Q7XHV2	Q7XHV2 oryza sativ
24	40	50.6	123	3	Q9PBH0	Q9PBH0 yarrowia li
25	40	50.6	152	16	Q9CN19	Q9CN19 pasteurella
26	40	50.6	154	16	Q8DB45	Q8DB45 vibrio vuln
27	40	50.6	205	2	Q52806	Q52806 amycolatops
28	40	50.6	212	2	Q54198	Q54198 streptomyce
29	40	50.6	377	2	Q9X5C1	Q9X5C1 escherichia
30	40	50.6	377	2	Q9X5C0	Q9X5C0 escherichia
31	40	50.6	377	16	Q8XDQ2	Q8XDQ2 escherichia
32	40	50.6	380	6	Q9TV3	Q9TV3 bos taurus
33	40	50.6	381	2	Q84G12	Q84G12 enterobacte
34	40	50.6	427	5	Q19368	Q19368 caenorhabdi
35	40	50.6	427	16	Q9CLD3	Q9CLD3 pasteurella
36	40	50.6	434	5	Q9U2F5	Q9U2F5 caenorhabdi
37	40	50.6	483	10	Q7XSV9	Q7XSV9 oryza sativ
38	40	50.6	524	5	Q8XU7	Q8XU7 diatrophila
39	40	50.6	541	17	Q97BK4	Q97BK4 thermoplasma
40	40	50.6	551	17	Q8TYW6	Q8TYW6 methanopyru
41	40	50.6	586	10	Q851F9	Q851F9 oryza sativ
42	40	50.6	623	10	Q9ZVN7	Q9ZVN7 arabidopsis
43	40	50.6	625	12	Q90881	Q90881 myxoma viru
44	40	50.6	659	16	Q67656	Q67656 aquifex aeo
45	40	50.6	746	11	Q9JL74	Q9JL74 mus musculu

## ALIGNMENTS

RESULT 1  
ID Q94257 PRELIMINARY; PRT; 117 AA.

AC Q94257; 01-FEB-1997 (TREMBLrel. 02, Created)  
DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)  
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
DE Hypothetical protein.  
GN K04A8.3.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
OC Rhabditidae; Pelodierinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RX MEDLINE=99069613; PubMed=9851916;  
RA None;  
RT "Genome sequence of the nematode C. elegans: a platform for  
investigating biology. The C. elegans Sequencing Consortium. ";  
RT Science 282:2012-2018(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RA Geisel C., Bradshaw H.;  
RT "The sequence of C. elegans cosmid K04A8. ";  
RL Submitted (NOV-1996) to the EMBL/Genbank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RA Waterston R.;  
RT "Direct Submission. ";  
RL Submitted (AUG-2001) to the EMBL/Genbank/DBJ databases.  
DR EMBL: U64849; AAC48049.1. -  
DR PIR: D89075; D89075.  
DR WormBase: K04A8.3; CE11704.  
KW Hypothetical protein.  
SQ  
SEQUENCE 117 AA; 12672 MW; 4749892B70E08919 CRC64;

Query Match 58.2%; Score 46; DB 5; Length 117;  
 Best Local Similarity 63.6%; Pred. No. 3.5;  
 Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 4 PAGSSNYVYG 14  
 |||||  
 DB 86 PGGSSQYLYGF 96

## RESULT 2

O89YMS PRELIMINARY; PRT; 307 AA.

AC O89YMS; 01-JUN-2003 (TREMBlrel. 24, Created)  
 DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)  
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
 DE Putative anti-sigma factor.  
 GN B74706.  
 OS Bacteroides thetaiotaomicron.  
 OC Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;  
 OC Bacteroidaceae; Bacteroides.  
 OK NCBI\_Taxid=818;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=VPI-5482 / ATCC 29148;  
 RX MEDLINE=22550858; PubMed=12663928;  
 RA Xu J., Bjursell M.K., Himrod J., Deng S., Carmichael L.K.,  
 RT "A genomic view of the human-bacteroides thetaiotaomicron symbiosis."  
 RL Science 299:2074-2076(2003).  
 DR EMBL; AE016946; AAC79811.1; -  
 DR GO; GO:0019013; C:viral nucleocapsid; IEA.  
 DR InterPro; IPR003486; Nairo nucleocap.  
 DR ProDom; PD006459; Nairo\_nucleocap; 1.  
 KW Complete proteome.

Query Match 57.0%; Score 45; DB 16; Length 307;  
 Best Local Similarity 81.8%; Pred. No. 16;  
 Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 3 TPAGSSNYVYG 13  
 |||||  
 DB 147 TPAGLSVTVYG 157

## RESULT 3

O8S6N2 PRELIMINARY; PRT; 1960 AA.

AC O8S6N2; 01-JUN-2002 (TREMBlrel. 21, Created)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)  
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
 DE Hypothetical protein.  
 GN OSJNBA0073101.15.  
 OS Oryza sativa (japonica cultivar-group).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoideae; Oryzaceae; Oryza.  
 OK NCBI\_Taxid=39947;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Nipponbare;  
 RA Buell C.R., Yuan Q., Qiyang S., Liu J., Gansberger K., Kim M.M.,  
 RA Overton II L.L., Bera J.J., Taitlin T., Kiol M.I., Dairahi B.B.,  
 RA Jin S.S., Koo H., Zismann V., Hsiao J., Blunt S., Varakan S.S.,  
 RA Uteberck T.T., Feldblum T.V., Yang Q.Q., Haas B.J., Suh B.B.,  
 RA Peterson J.J., Quackenbush J., White O., Salzberg S.L., Fraser C.M.,  
 RT "Oryza sativa chromosome 10 BAC OSJNBA0073101 genomic sequence."  
 RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.

RA The Rice Chromosome 10 Sequencing Consortium,  
 RT "In-depth view of structure, activity, and evolution of rice  
 RT chromosome 10."  
 RL Science 300:1566-1569(2003).

RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Nipponbare;  
 RA Buell C.R., Wang R.A., McCombie W.R., Messing I., Yuan Q.,  
 RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AC092548; AAM18735.1; -  
 DR EMBL; AE017083; AAP53324.1; -  
 DR Gramene; O8S6N2; -  
 KW Hypothetical protein.

Query Match 57.0%; Score 45; DB 10; Length 1960;  
 Best Local Similarity 42.9%; Pred. No. 1.3e+02;  
 Matches 6; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

OY 1 HTTPAGSSNYVYG 14  
 |||||  
 DB 325 HLPQSSNEWYIGW 338

## RESULT 4

O90TC3 PRELIMINARY; PRT; 838 AA.

AC O90TC3; 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
 DT 01-MAY-2002 (TREMBlrel. 20, Last annotation update)  
 DE UL22 product homolog (Glycoprotein H).  
 GN ORF 29 OR UL22.  
 OS Marek's disease virus serotype 2 MDV2, and  
 OS Gallid herpesvirus 3  
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
 OC Alphaherpesvirinae; Marek's disease-like viruses.  
 OK NCBI\_Taxid=36353, 35250;  
 RN [1]

RP SEQUENCE FROM N.A.  
 RC SPECIES-Gallid herpesvirus 1 (serotype 2); STRAIN=HPRS24;  
 RA Jang H., Cai J., Izumiya Y., Murakami Y., Mochizuki M., Song C.,  
 RA Lee Y., Kai C., Takahashi E., Mikami T.,  
 RT "The complete DNA sequence and transcription map of the unique long  
 RT genome region of Marek's disease virus type 2."  
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.

RP SEQUENCE FROM N.A.  
 RC SPECIES-Gallid herpesvirus 3; STRAIN=HPRS24;  
 RA Izumiya Y., Jang H., Ono M., Mikami T.,  
 RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.

RP SEQUENCE FROM N.A.  
 RC SPECIES-Gallid herpesvirus 3; STRAIN=HPRS24;  
 RA Izumiya Y., Jang H., Ono M., Mikami T.,  
 RT "A Complete Genomic DNA Sequence of Marek's Disease Virus Type 2,  
 RT Strain HPRS24."  
 RL Curr. Top. Microbiol. Immunol. 0:0-0(2000).  
 DR EMBL; AB024414; BAA82917.1; -  
 DR EMBL; AB049735; BAB16531.1; -  
 DR InterPro; IPR003493; Herpes\_glycopH.  
 DR Pfam; PR02489; Herpes\_glycop\_H; 1

Query Match 55.7%; Score 44; DB 12; Length 838;  
 Best Local Similarity 34.6%; Pred. No. 73;  
 Matches 9; Conservative 4; Mismatches 1; Indels 12; Gaps 1;

OY 1 HT-----PAGSSNYVYG 14  
 |||||  
 DB 327 HITQASHRLFTISAMPASTSYNHFP 352

```

RESULT 5
Q859Q1 ID Q859Q1 PRELIMINARY; PRT; 150 AA.
AC Q859Q1;
DT 01-JUN-2003 (TREMBlrel. 24, Created)
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE 16 kDa protein.
OS Bacteriophage N4.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Podoviridae.
OX NCBI_TaxID=10752;
RN [1]
RP SEQUENCE FROM N.A.
RA Kazmierczak K.M., Kottman-Denes L.B.;
RT "Partial sequence and sequence analysis of the Bacteriophage N4 virion
RT RNA polymerase and five other phage late gene products."
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF350004; A024829.1; -
SQ SEQUENCE 150 AA; 16094 MW; 59F013A0809A600 CRC64;

Query Match 54.4%; Score 43; DB 9; Length 150;
Best Local Similarity 58.3%; Pred. No. 15;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 3 TPAGSSNYVYGY 14
Db 9 TPAGSSNYVYGY 20

RESULT 6
Q33707 ID Q33707 PRELIMINARY; PRT; 208 AA.
AC Q33707;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Putative epimerase.
GN DNMU.
OS Streptomyces puceatus.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomycetes.
OX NCBI_TaxID=1950;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=ATCC 29050;
RX MEDLINE=97352712; PubMed=9209071;
RA Otten S.L., Gallo M.A., Madduri K., Liu X., Hutchinson C.R.;
RT "Cloning and characterization of the Streptomyces puceatus dnmzuv
RT genes encoding three enzymes required for biosynthesis of the
RT daunosubicin precursor thymidine diphospho-L-daunosamine."
RL J. Bacteriol. 179:4446-4450(1997).
DR EMBL; AF006633; AA663046.1; -
DR HSSP; P26394; IDZR.
DR GO; GO:0003824; P: catalytic activity; IEA.
DR GO; GO:0009103; P: lipidopolysaccharide biosynthesis; IEA.
DR InterPro; IPR000888; dtdp_sugar_isom.
DR Pfam; PF00908; dtdp_sugar_isom; 1.
DR ProDom; PD001462; dtdp_sugar_isom; 1.
SQ SEQUENCE 208 AA; 22632 MW; 5B6C186853A8529 CRC64;

Query Match 54.4%; Score 43; DB 2; Length 208;
Best Local Similarity 63.6%; Pred. No. 22;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 ITPAGSSNYVY 12
Db 66 VTPGSAKYVY 76

RESULT 7
Q9ZUA7 ID Q9ZUA7 PRELIMINARY; PRT; 269 AA.
AC Q9ZUA7;

```

```

DT 01-MAY-1999 (TREMBlrel. 10, Created)
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE At2g01790 protein.
GN AT2G01790.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RP STRAIN=cv Columbia;
RX MEDLINE=20083487; PubMed=10617197;
RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.B., Feldblum T.V.,
RA Buell C.R., Ketchum K.A., Lee J.U., Ronning C.M., Koo H., Moffat K.S.,
RA Cronin L.A., Shen M., Vanaken S.E., Umayam L., Tallon L.J., Gill J.E.,
RA Adams M.D., Carrera A.U., Creasy T.H., Goodman H.M., Somerville C.R.,
RA Copenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,
RA Salzberg S.L., Fraser C.M., Venter J.C.;
RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis
RT thaliana."
RL Nature 402:761-768(1999).
RN [2]
RP SEQUENCE FROM N.A.
RP STRAIN=cv Columbia;
RA Lin X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC006069; AAD12693.1; -
DR PIR; B84429; B84429.
DR InterPro; IPR002083; MATH.
DR InterPro; IPR008974; Traf_dom.
DR Pfam; PF00917; MATH; 1.
DR SMART; SM00061; MATH; 1.
DR PROSITE; PSS0144; MATH; 1.
SQ SEQUENCE 269 AA; 30515 MW; C5D3BDBA95E32B84 CRC64;

Query Match 54.4%; Score 43; DB 10; Length 269;
Best Local Similarity 54.5%; Pred. No. 30;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 4 PAGSSNYVYGY 14
Db 42 PEGNNNYLYDY 52

RESULT 8
Q8FWT3 ID Q8FWT3 PRELIMINARY; PRT; 402 AA.
AC Q8FWT3;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Site-specific recombinase, phage integrase family.
GN BRA0362.
OS Brucella suis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Brucellaceae; Brucella.
OX NCBI_TaxID=29461;
RN [1]
RP SEQUENCE FROM N.A.
RP STRAIN=1330 / B1ovar 1;
RX MEDLINE=22247741; PubMed=12271122;
RA Paulsen I.T., Seshadri R., Nelson K.E., Eisen J.A., Heidelberg J.F.,
RA Read T.D., Dodson R.J., Umayam L., Brinkac L.M., Beaman M.J.,
RA Daugherty S.C., Debey R.T., Durkin A.S., Kolonay J.F., Madupu R.,
RA Nelson W.C., Ayodeji B., Kraul M., Shetty J., Malik J., Van Aken S.E.,
RA Riedmuller S., Tettelin H., Gill S.R., White O., Salzberg S.L.,
RA Hoover D.L., Lindler L.E., Hailing S.W., Boyle S.W., Fraser C.M.;
RT "The Brucella suis genome reveals fundamental similarities between
RT animal and plant pathogens and symbionts."
RL Proc. Natl. Acad. Sci. U.S.A. 99:13148-13153(2002).

```

DR EMBL; AB014535; AAN33560.1; -.  
 DR TIGR; BR0362; -.  
 DR GO; GO:0003677; F:DNA binding; IEA.  
 DR GO; GO:0015074; P:DNA integration; IEA.  
 DR GO; GO:0006310; P:DNA recombination; IEA.  
 DR InterPro; IPR02104; Phage integrase.  
 DR Pfam; PF00589; Phage\_integrase; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 402 AA; 46265 MW; FE18D1FEB524EAB3 CRC64;

Query Match 54.4%; Score 43; DB 16; Length 402;  
 Best Local Similarity 61.5%; Pred. No. 47;  
 Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 2 ITPAGSSNYVYGY 14  
 |||||::||  
 Db 29 ITPAGSISFRVY 41

RESULT 9  
 ID Q82410 PRELIMINARY; PRT; 337 AA.

AC Q82410; 01-JUN-2003 (TRENBLrel. 24, Created)  
 DT 01-JUN-2003 (TRENBLrel. 24, Last sequence update)  
 DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)  
 DE Peptide chain release factor 2.  
 GN PRFB OR CCA00166.  
 OS Chlamydomonas caviae.  
 OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.  
 OX NCBI\_TaxID=83557;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=GPIC;  
 RX MEDLINE=22569155; PubMed=12682364;  
 RA Read T.D., Myers G.S.A., Brumham R.C., Nelson W.C., Paulsen I.T.,  
 RA Heidelberg J., Holtzapple E., Khouri H., Fedorova N.B., Carty H.A.,  
 RA Umayam L.A., Haft D.H., Peterson J., Beaman M.J., White O.,  
 RA Salzberg S.L., Hsia R.-C., McClarty G., Rank R.G., Bavoil P.M.,  
 RA Fraser C.M.;  
 RT "Genome sequence of Chlamydomonas caviae (Chlamydia psittaci GPIC):  
 RT examining the role of niche-specific genes in the evolution of the  
 RT Chlamydiaceae.";  
 RL Nucleic Acids Res. 31:2134-2147(2003).  
 DR EMBL; AE016994; AAP04917.1; -.  
 DR TIGR; CCA00166; -.  
 DR GO; GO:0005737; C:cytoplasm; IEA.  
 DR GO; GO:0016149; F:translational release factor activity, codon . . .; IEA.  
 DR GO; GO:0006415; P:translational termination; IEA.  
 DR InterPro; IPR005139; PCRF.  
 DR InterPro; IPR000352; PEP\_rel\_factor\_1.  
 DR InterPro; IPR004374; PRFB.  
 DR Pfam; PF03462; PCRF; 1.  
 DR Pfam; PF00472; RF-1; 1.  
 DR TIGRPFAM; TIGR00020; PRFB; 1.  
 DR PROSITE; PS00745; RF\_PROK\_1; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 337 AA; 38702 MW; 03B441E6176B1FF CRC64;

Query Match 53.2%; Score 42; DB 16; Length 337;  
 Best Local Similarity 57.1%; Pred. No. 58;  
 Matches 8; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 ITPAGSSNYVYGY 14  
 |||||::||  
 Db 145 HITVKGSEYAYGY 158

RESULT 10  
 ID Q7UB23 PRELIMINARY; PRT; 400 AA.  
 AC Q7UB23; 01-OCT-2003 (TRENBLrel. 25, Created)  
 DT 01-OCT-2003 (TRENBLrel. 25, Created)

DT 01-OCT-2003 (TRENBLrel. 25, Last sequence update)  
 DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)  
 DE Putative integrase.  
 GN S2723.  
 OS Shigella flexneri.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Shigella.  
 OX NCBI\_TaxID=623;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=2457T / ATCC 700930 / Serotype 2a;  
 RX MEDLINE=22590274; PubMed=12704152;  
 RA Wei J., Goldberg M.B., Burland V., Venkatesan M.M., Deng W.,  
 RA Fournier G., Mayhew G.F., Plunkett G. III, Rose D.J., Darling A.,  
 RA Mau B., Perna N.T., Payne S.M., Runyan-Joneck L.J., Zhou S.,  
 RA Schwartz D.C., Blattner F.R.;  
 RT "Complete genome sequence and comparative genomics of Shigella  
 RT flexneri serotype 2a strain 2457T.";  
 RL Infect. Immun. 71:2775-2786(2003).  
 DR EMBL; AB016986; AAP17879.1; -.  
 SQ SEQUENCE 400 AA; 45329 MW; 0EC6ADP2527FEBD5 CRC64;

Query Match 53.2%; Score 42; DB 16; Length 400;  
 Best Local Similarity 53.8%; Pred. No. 70;  
 Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 2 ITPAGSSNYVYGY 14  
 :|||::||  
 Db 29 VTPAGSISFRVY 41

RESULT 11  
 ID Q9RM70 PRELIMINARY; PRT; 505 AA.

AC Q9RM70; 01-MAY-2000 (TRENBLrel. 13, Created)  
 DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)  
 DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)  
 DE Putative membrane protein CjaE.  
 GN CjaE.  
 OS Campylobacter jejuni.  
 OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;  
 OC Campylobacteraceae; Campylobacter.  
 OX NCBI\_TaxID=197;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=72Dz/92;  
 RA Pawelec D.;  
 RT Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AJ249744; CAB56502.1; -.  
 DR GO; GO:0016020; C:membrane; IEA.  
 DR GO; GO:0015035; F:protein disulfide oxidoreductase activity; IEA.  
 DR GO; GO:0006118; P:electron transport; IEA.  
 DR InterPro; IPR003752; Dsdb.  
 DR Pfam; PF02600; Dsdb; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 505 AA; 56645 MW; 10A8E12EA925F826 CRC64;

Query Match 53.2%; Score 42; DB 2; Length 505;  
 Best Local Similarity 46.2%; Pred. No. 92;  
 Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Qy 2 ITPAGSSNYVYGY 14  
 :|||::||  
 Db 23 LIPVGIANFVFGY 35

RESULT 12  
 ID Q960R6 PRELIMINARY; PRT; 661 AA.  
 AC Q960R6; 01-DEC-2001 (TRENBLrel. 19, Created)  
 DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)  
 DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)







RP SEQUENCE FROM N.A.  
 RC SPECIES=S.pyogenes; STRAIN=SSI-1 / Serotype M3;  
 RA Nakagawa I., Kurokawa K., Nakata M., Tomiyasu Y., Yamashita A.,  
 RA Yamazaki K., Okahashi N., Kawabata S., Yasunaga T., Hattori M.,  
 RA Hayaishi H., Hamada S.,  
 RT "The genome of invasive Streptococcus pyogenes; a comparative analysis  
 of S. pyogenes SSI-1, SF370 and MGAS8232."  
 RL Submitted (May-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AE006583; AAK34295.1; -  
 DR EMBL; AE010066; AAL98085.1; -  
 DR EMBL; AE014159; AAM79759.1; -  
 DR EMBL; AP005143; BAC3805.1; -  
 DR HSP; O31408; IBA.  
 DR GO; GO:0003700; P:transcription factor activity; IEA.  
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.  
 DR InterPro; IPR01669; Arg\_repress.  
 DR Pfam; PF01316; Arg\_repressor; 1.  
 DR PRINTS; PRO1467; ARGREPRESSOR.  
 DR ProDom; PD007402; Arg\_repress; 1.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 156 AA; 17454 MW; CD086D497A3027C4 CRC64;

Query Match 51.9%; Score 41; DB 16; Length 156;  
 Best Local Similarity 60.0%; Pred. No. 36;  
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 4 PAGSSNYVG 13  
 :|||:  
 :|||:  
 Db 54 PGGGGRYVG 63

## RESULT 15

O8ZUN9 PRELIMINARY; PRT; 306 AA.  
 AC O8ZUN9;  
 DT 01-MAR-2002 (TREMblrel. 20, Created)  
 DT 01-MAR-2002 (TREMblrel. 20, Last sequence update)  
 DT 01-JUN-2003 (TREMblrel. 24, Last annotation update)  
 DE 2-hydroxyhepta-2,4-diene-1,7-dioate isomerase, conjectural.  
 GN PAE2688.  
 OS Pyrobaculum aerophilum.  
 OC Archaea; Crenarchaeota; Thermoprotei; Thermoproteales;  
 OC Thermoproteaceae; Pyrobaculum.  
 OX NCBI\_TaxID=13773;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=IM2 / ATCC 51768 / DSM 7523;  
 RX MEDLINE=21664397; PubMed=11792869;  
 RA Fitz-Gibbon S.T., Ladner H., Kim U.-J., Stetter K.O., Simon M.I.,  
 RA Miller J.H.,  
 RT "Genome sequence of the hyperthermophilic crenarchaeon Pyrobaculum  
 aerophilum".  
 RT Proc. Natl. Acad. Sci. U.S.A. 99:984-989(2002).  
 RL EMBL; AE009889; AAL64367.1; -  
 DR EMBL; AE009889; AAL64367.1; -  
 DR GO; GO:0016853; P:isomerase activity; IEA.  
 DR GO; GO:0008152; P:metabolism; IEA.  
 DR InterPro; IPR02529; FAA\_hydrolase.  
 DR Pfam; PF01557; FAA\_hydrolase; 1.  
 KW Isomerase; Complete proteome.  
 SQ SEQUENCE 306 AA; 34384 MW; 137048EEB3B97347 CRC64;

Query Match 51.9%; Score 41; DB 17; Length 306;  
 Best Local Similarity 50.0%; Pred. No. 77;  
 Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

OY 1 HITPAGSSNYVG 14  
 :|||:  
 :|||:  
 Db 153 YIDPARADIVFGY 166

Search completed: October 5, 2004, 08:13:38  
 Job time : 11.4105 secs

This Page Blank (uspto)



XX Sequence 14 AA;  
SQ

Query Match 100.0%; Score 79; DB 4; Length 14;  
Best Local Similarity 100.0%; Pred. No. 8,4e-07;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HITPAGSSNYYVGY 14  
Db 1 HITPAGSSNYYVGY 14

RESULT 2  
AAE10555  
ID AAE10555 standard; peptide; 131 AA.  
XX AAE10555;  
AC AAE10555;  
XX  
DT 10-DEC-2001 (first entry)  
XX  
DE HPL inhibiting VHH fragment, HPL #15 from llama species.  
XX  
KM llama antibody; camelid; anorectic; heavy chain variable domain; VHH;  
KM human dietary enzyme inhibitor; medicament; human pancreatic lipase; HPL;  
KM food; human gastric lipase; HGL; cosmetic control; body weight.  
XX  
OS llama sp.  
XX  
FH Key  
FT Location/Qualifiers  
FT 31..35  
FT /label= CDR1  
FT /note= "Complementarity determining region 1"  
FT 50..64  
FT /label= CDR2  
FT /note= "Complementarity determining region 2"  
FT 98..111  
FT /label= CDR3  
FT /note= "Complementarity determining region 3"  
XX  
FN EP1134231-A1.  
XX  
PD 19-SEP-2001.  
XX  
PF 20-FEB-2001; 2001EP-00200703.  
XX  
PR 14-MAR-2000; 2000EP-00200930.  
XX  
PA (UNIL ) UNILEVER NV.  
PA (UNIL ) UNILEVER PLC.  
XX  
PI Bezemer S, Van De Burg M, De Haard JW, Tarellus E;  
XX  
DR WPI; 2001-572718/65.  
XX  
PT New antibody or its fragments for inhibiting human dietary enzymes,  
PT useful for cosmetic control of body weight of human beings, comprises  
PT heavy chain variable domain derived from immunoglobulin naturally devoid  
PT of light chains.  
XX  
PS Example 2; Page 9; 37pp; English.  
XX  
CC The patent discloses antibodies or their fragments comprising a heavy  
CC chain variable domain (VHH) derived from an immunoglobulin naturally  
CC devoid of light chains specific for inhibiting human dietary enzymes. The  
CC antibodies of the invention are useful for the preparation of medicaments  
CC or food for inhibiting the activity of one or more human dietary enzymes  
CC especially human pancreatic lipase (HPL) or human gastric lipase (HGL).  
CC which are useful for the cosmetic control of body weight of human beings.  
CC The present peptide sequence is HPL inhibiting VHH fragment, HPL #15 from  
CC llama (camelid) species  
XX  
SQ Sequence 131 AA;

Query Match 100.0%; Score 79; DB 4; Length 131;  
Best Local Similarity 100.0%; Pred. No. 1.1e-05;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HITPAGSSNYYVGY 14  
Db 98 HITPAGSSNYYVGY 111

RESULT 3  
ABU97103  
ID ABU97103 standard; protein; 110 AA.  
XX  
AC ABU97103;  
XX  
DT 26-JUL-2003 (first entry)  
XX  
DE Recombinant anti-IBDV-antibody light chain variable region protein #31.  
XX  
KM Recombinant antibody; IBDV; infectious bursal disease virus; virucide;  
KM gene therapy; heavy chain variable region; light chain variable region;  
KM immunodiagnostic; immunotherapeutic; infection.  
XX  
OS Gallus gallus.  
XX  
OS WO200298921-A1.  
XX  
PN 12-DEC-2002.  
XX  
PD 05-JUN-2002; 2002MO-AU000729.  
XX  
PF 05-JUN-2001; 2001AU-00005468.  
XX  
PR (CSIR ) COMMONWEALTH SCI & IND RES ORG.  
XX  
PI Ignjatovic J, Prowse S, Heine HG, Sapats SI;  
XX  
DR WPI; 2003-167324/16.  
XX  
DR N-PSDB; ACA9348.  
XX  
XX  
PT New recombinant antibody fragment having a binding specificity for an  
PT infectious bursal disease virus (IBDV) antigenic determinant, useful for  
PT the prevention, diagnosis and treatment of IBDV related ailments in  
PT animals.  
XX  
PS Claim 4; Fig 4I; 265pp; English.  
XX  
CC The invention describes a new recombinant antibody fragment (II) with  
CC specificity for an infectious bursal disease virus (IBDV) antigenic  
CC determinant, comprises at least a variable region having a heavy chain  
CC region (VH) with any of 35 amino acid sequences, and a light (VL) chain  
CC region with any of 42 amino acid sequences given in the specification  
CC where (I) additionally comprises any of 42 sequences amino acid  
CC sequences. The process is useful for identifying recombinant IBDV  
CC antibodies as immunodiagnostic or immunotherapeutic agents. (II) is useful  
CC for the diagnosis of an infection caused by at least one IBDV strain. The  
CC other methods and compositions of the present invention are useful for  
CC the prevention, diagnosis and treatment against IBDV related ailments in  
CC animals. This is the amino acid sequence of a recombinant anti-IBDV-  
CC antibody light chain variable region  
XX  
SQ Sequence 110 AA;

Query Match 55.7%; Score 44; DB 6; Length 110;  
Best Local Similarity 61.5%; Pred. No. 11;  
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 ITPAGSSNYYVGY 14  
Db 19 ITCGSGSNYYVGM 31

RESULT 4

ABU97120  
ID ABU97120 standard; protein; 250 AA.  
XX  
AC ABU97120;  
XX  
DT 28-JUL-2003 (first entry)  
XX  
DE Recombinant anti-infective bursal disease virus-antibody fragment #6.  
XX  
KM Recombinant antibody; IBDV; infectious bursal disease virus; virucide;  
KM gene therapy; heavy chain variable region; light chain variable region;  
KM immunodiagnostic; immunotherapeutic; infection.  
XX  
OS Gallus gallus.  
XX  
PN WO200298921-A1.  
XX  
PD 12-DEC-2002.  
XX  
PF 05-JUN-2002; 2002WO-AU000729.  
XX  
PR 05-JUN-2001; 2001AU-00005468.  
XX  
PA (CSIR ) COMMONWEALTH SCI & IND RES ORG.  
XX  
PI Ignjatovic J, Prowse S, Heine HG, Sepats SI;  
XX  
DR WPI; 2003-167324/16.  
XX  
PT New recombinant antibody fragment having a binding specificity for an  
PT infectious bursal disease virus (IBDV) antigenic determinant; useful for  
PT the prevention, diagnosis and treatment of IBDV related ailments in  
PT animals.  
XX  
PS Claim 6; Page 211-212; 265pp; English.  
XX  
CC The invention describes a new recombinant antibody fragment (I) with  
CC specificity for an infectious bursal disease virus (IBDV) antigenic  
CC determinant, comprises at least a variable region having a heavy chain  
CC region (VH) with any of 35 amino acid sequences, and a light (VL) chain  
CC region with any of 42 amino acid sequences given in the specification  
CC where (I) additionally comprises any of 42 sequences amino acid  
CC sequences. The process is useful for identifying recombinant IBDV  
CC antibodies as immunodiagnostic or immunotherapeutic agents. (I) is useful  
CC for the diagnosis of an infection caused by at least one IBDV strain. The  
CC other methods and compositions of the present invention are useful for  
CC the prevention, diagnosis and treatment against IBDV related ailments in  
CC animals. This is the amino acid sequence of a recombinant anti-IBDV-  
CC antibody fragment  
XX  
SQ Sequence 250 AA;  
XX  
Query Match 55.7%; Score 44; DB 6; Length 250;  
Best Local Similarity 61.5%; Pred. No. 29;  
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;  
XX  
OY 2 ITPAGSSNYVGY 14  
DB 159 ITCGSGSNYYIGW 171  
XX  
RESULT 5  
AAW75775  
ID AAW75775 standard; protein; 633 AA.  
XX  
AC AAW75775;  
XX  
DT 02-DEC-1998 (first entry)  
XX  
DE Amino acid sequence of lepidopteran-active 8612 toxin.  
XX  
KM 8612 toxin; PCR; primer; amplification; Bacillus thuringiensis; probe;  
KM lepidoptera; pest; pesticide; Ostrinia nubilalis; Heliothis virescens;

KM Helicoverpa zea; hybridisation.  
XX  
OS Bacillus thuringiensis.  
XX  
PN WO9840490-A1.  
XX  
PD 17-SEP-1998.  
XX  
PF 13-MAR-1998; 98WO-US005081.  
XX  
PR 13-MAR-1997; 97US-0040512P.  
XX  
PA (MYCO ) MYCOGEN CORP.  
XX  
PI Schnepf HE, Narva KE, Muller-Cohn J;  
XX  
DR WPI; 1998-506734/43.  
DR N-PSDB; AAV52612.  
XX  
PT New insecticidal Bacillus thuringiensis toxins - useful for controlling  
PT lepidopteran pests, especially Ostrinia nubilalis, Heliothis virescens  
PT and Helicoverpa zea.  
XX  
PS Claim 5; Page 36-38; 50pp; English.  
XX  
CC This is the amino acid sequence of a novel Bacillus thuringiensis toxin  
CC used in the method of the invention, to control lepidopteran pests. The  
CC new toxins are useful as pesticides, especially for the control of  
CC Ostrinia nubilalis, Heliothis virescens, and Helicoverpa zea. The  
CC polynucleotide coding sequences are useful for recombinant expression of  
CC the toxins and the primers, together with probes derived from the new  
CC sequences, are useful for the identification and characterisation of  
CC novel genes that encode pesticidal toxins  
XX  
SQ Sequence 633 AA;  
XX  
Query Match 55.7%; Score 44; DB 2; Length 633;  
Best Local Similarity 66.7%; Pred. No. 83;  
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
XX  
OY 3 TPAGSSNYVGY 14  
DB 410 TPRGNSNYVGY 421  
XX  
RESULT 6  
AAW75774  
ID AAW75774 standard; protein; 633 AA.  
XX  
AC AAW75774;  
XX  
DT 02-DEC-1998 (first entry)  
XX  
DE Amino acid sequence of lepidopteran-active HD525 toxin.  
XX  
KM HD525 toxin; PCR; primer; amplification; Bacillus thuringiensis; probe;  
KM lepidoptera; pest; pesticide; Ostrinia nubilalis; Heliothis virescens;  
KM Helicoverpa zea; hybridisation.  
XX  
OS Bacillus thuringiensis.  
XX  
PN WO9840490-A1.  
XX  
PD 17-SEP-1998.  
XX  
PF 13-MAR-1998; 98WO-US005081.  
XX  
PR 13-MAR-1997; 97US-0040512P.  
XX  
PA (MYCO ) MYCOGEN CORP.  
XX  
PI Schnepf HE, Narva KE, Muller-Cohn J;

DR MPI, 1998-506734/43.  
 DR N-PSDB; AAV52611.  
 XX New insecticidal *Bacillus thuringiensis* toxins - useful for controlling  
 PT lepidopteran pests, especially *Ostrinia nubilalis*, *Heliothis virescens*  
 PT and *Helicoverpa zea*.  
 XX  
 PS Claim 14; Page 32-34; 50pp; English.  
 XX  
 CC This is the amino acid sequence of a novel *Bacillus thuringiensis* toxin  
 CC used in the method of the invention, to control lepidopteran pests. The  
 CC new toxins are useful as pesticides, especially for the control of  
 CC *Ostrinia nubilalis*, *Heliothis virescens*, and *Helicoverpa zea*. The  
 CC polynucleotide coding sequences are useful for recombinant expression of  
 CC the toxins and the primers, together with probes derived from the new  
 CC sequences, are useful for the identification and characterisation of  
 CC novel genes that encode pesticidal toxins  
 XX  
 SQ Sequence 633 AA;  
 CC  
 CC Query Match 55.7%; Score 44; DB 2; Length 633;  
 CC Best Local Similarity 66.7%; Pred. No. 83;  
 CC Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
 QY 3 TPAGSSNYVGY 14  
 Db 410 TPRGSSNYYPGY 421  
 CC  
 CC RESULT 7  
 CC ABB71017  
 CC ID ABB71017 standard; protein; 128 AA.  
 CC XX  
 CC AC ABB71017;  
 CC XX  
 CC DT 26-MAR-2002 (first entry)  
 CC XX  
 CC DE *Drosophila melanogaster* polypeptide SEQ ID NO 39843.  
 CC XX  
 CC KW *Drosophila*; developmental biology; cell signalling; insecticide;  
 CC XX pharmaceutical.  
 CC OS *Drosophila melanogaster*.  
 CC XX  
 CC PN WO200171042-A2.  
 CC PD 27-SEP-2001.  
 CC XX  
 CC PF 23-MAR-2001; 2001WO-US009221.  
 CC XX  
 CC PR 23-MAR-2000; 2000US-0191637P.  
 CC XX 11-JUL-2000; 2000US-00614150.  
 CC XX  
 CC PA (PEKE ) PE CORP NY.  
 CC XX  
 CC PI Venter JC, Adams M, Li PWD, Myers EW;  
 CC DR WP; 2001-656860/75.  
 CC DR N-PSDB; ABL15120.  
 CC XX  
 CC PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 CC PT genes from *Drosophila* and for elucidating cell signaling and cell-cell  
 CC PT interactions.  
 CC  
 CC Disclosure; SEQ ID NO 39843; 21pp + Sequence Listing; English.  
 CC  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from *Drosophila*. The invention is  
 CC useful in developmental biology and in elucidating cell signaling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 CC sequences (ABL01840-ABL16175) and the encoded proteins (ABBS7737-

CC	ABB720722)	The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pat_sequences
XX	Sequence 128 AA:	
SQ	Query Match	54.4%; Score 43; DB 4; Length 128;
	Best Local Similarity	53.8%; Pred. No. 20;
	Matches	7; Conservative 3; Mismatches 3; Indels 0; Gaps 0
OY	2 ITPAGSSNTVYG 14 ::   ::    ::   ::	
D6	14 LTPGSSSTSVYGY 26	
RESULT 8		
ABP43116	ID ABP43116 standard; protein; 175 AA.	
XX	ABP43116;	
XX	22-AUG-2002 (first entry)	
DE	Human ovarian antigen HVAHAJ23, SEQ ID NO:4248.	
KX	Human; ovarian antigen; ovary; ovarian; breast; cancer; tumour;	
KW	ovarian cancer; breast cancer; tumour; reproductive system disorder;	
KW	fertility; pregnancy disorder; anovulation; polycystic ovary syndrome;	
KW	PCOS; ovarian cyst; dysmenorrhea; endocrine disorder; infection;	
KW	inflammatory condition; immune disorder; blood disorder;	
KW	cardiovascular disorder; respiratory disorder; neurological disorder;	
KW	gastrointestinal disorder; urinary system disorder; drug screening;	
KW	gene therapy; chromosome mapping; forensic analysis;	
KW	antibody preparation; cytostatic; immunomodulatory; neuroprotective;	
KW	antiinflammatory; gynaecological; reproductive.	
XX	Homo sapiens.	
OS	Homosapiens.	
PN	WO200200677-A1.	
XX	03-JAN-2002.	
PD	07-JUN-2001; 2001MO-US018559.	
PF	07-JUN-2001; 2000US-0209467P.	
PR	(HUMA-) HUMAN GENOME SCI INC.	
PA	Birse CE, Rosen CA;	
PI	WPI; 2002-147878/19.	
XX	N-PSDB; ABO56193.	
DR	Isolated nucleic acid molecules encoding novel ovarian polypeptides,	
XX	useful in the prevention, treatment and diagnosis of cancer (e.g. ovarian	
PT	cancer), immune disorders, cardiovascular disorders and neurological	
PT	diseases.	
Claim 11; SEQ ID NO 4248; 2922bp; English.		
XX	The invention relates to 2175 novel human ovarian antigens (ABP41054-	
CC	ABP43228) and to cDNAs encoding them (ABO54131-ABO56305), and also	
CC	encompasses polypeptides 90% identical and polynucleotides 95% identical	
CC	to the sequences of the invention. The invention additionally relates to	
CC	recombinant vectors and host cells comprising human ovarian antigen	
CC	polynucleotides, antibodies against human ovarian antigens, and the use	
CC	of ovarian antigen polynucleotides and polypeptides in diagnosing,	
CC	treating, prognosing or preventing various ovary and/or breast-related	
CC	disorders. Such conditions include ovarian cancer and breast cancer, and	
CC	metastatic tumours of ovarian or breast origin, reproductive system	
CC	disorders (e.g., infertility, disorders of pregnancy, anovulation,	
CC	polycystic ovary syndrome, ovarian cysts, and dysmenorrhoea), endocrine	
CC	disorders, infections (e.g., chlamydia, HIV, toxoplasmosis), and toxic	



shock syndrome), inflammatory conditions (e.g. mastitis, oophoritis and vaginitis), immune disorders (e.g. congenital and acquired immunodeficiencies, autoimmune oophoritis, systemic lupus erythematosus), blood-related disorders (e.g., anaemia), cardiovascular disorders, respiratory disorders, neurological disorders, gastrointestinal disorders and urinary system disorders. Ovarian antigen polypeptides and polynucleotides may also be used in screening for compounds which modulate ovarian antigen expression or activity. The polynucleotides may further be used for gene therapy, chromosome mapping, in the identification of individuals and in forensic analysis, and the polypeptides may be used as food additives or to prepare antibodies useful in disease diagnosis, drug targeting and phenotyping. The present sequence represents a human ovarian antigen of the invention. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at [ftp.wipo.int/pub/published\\_pat\\_sequences](http://wipo.int/pub/published_pat_sequences)

SQ Sequence 175 AA;

Query Match	54.4%;	Score 43;	Length 175;
Best Local	Pred. 66.7%;	DB 5;	
Matches	8; Conservative	1; Mismatches	3; Indels 0; Gaps 0;

```
QY      3 TPAGSSNYVGY 14
          | | | | | : |
Db      20 TDVGSSNYVFWY 31
```

RESULT 9  
ABP11310  
ID ABP11310 standard; protein; 54 AA.

AC ABP11310;

DT 25-JUN-2002 (first entry)

Human ORFX protein sequence SEQ ID NO:22602.

KM Human; open reading frame; ORF; gene therapy cancer; cancer; cirrhosis;  
KM hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;  
KM degenerative disorder; osteoarthritis; neurodegenerative disorder;  
KM cardiovascular disease; diabetes mellitus; systemic lupus erythematosus  
KM hypertension; hypothyroidism; cholesterol ester storage disease;  
KM immune deficiency; immune disorder; infectious disease;  
KM autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;  
KM myasthenia gravis.

OS Homo sapiens.

PN W0200192523-A2.

PD 06-DEC-2001

PF 29-MAY-2001; 2001WO-US010836.

PR	30-MAY-2000;	2000US-0206132P.
DT	30-MAY-2000;	2000US-0206132P.

XX  
F2 (CITPA - ) CИПАСЕН СОВБ

XX		Shimkrets BA	Teach N
PI			

XX  
DB  
WPT: 2002-106308/14.

DR N-PSDB; ABN2/002.  
XX

novel radical polymer preventing and tri

PT hyperproliferative disorders and autoimmune disorders.

PS Disclosure; SEQ ID NO 22602; 1037pp; English.

The present invention describes substantially purified human proteins (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1

in the specification) AEN15762 to AEN27252 encode the human ORFX proteins given in ABP00101 to ABP11500. ORFX proteins are useful for treating or preventing a pathology associated with an ORFX-associated disorder in humans, and in the manufacture of a medicament for treating a syndrome associated with ORFX-associated disorder. ORFX polynucleotide sequences can be used in gene therapy. ORFX sequences can be used in the treatment of cancer, hyperproliferative disorders, cirrhosis of liver, psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage, osteoarthritis, neurodegenerative disorders, disorders related to organ transplantation, cardiovascular diseases, diabetes mellitus, systemic lupus erythematosus, hypertension, hypothyroidism, cholesterol ester storage disease, various immune deficiencies and disorders, infectious diseases, autoimmune disorders such as multiple sclerosis, rheumatoid arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host disease and autoimmune inflammatory eye disease. ORFX proteins are also useful for treating burns, incisions, ulcers, for treating osteoporosis, bone degenerative disorders, or periodontal disease, and for gut protection or regeneration and treatment of lung or liver fibrosis, reperfusion injury in various tissues and conditions resulting from systemic cytokine damage. N.B. The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at [ftp.wipo.int/pub/published\\_pct\\_sequences](http://ftp.wipo.int/pub/published_pct_sequences)

Sequence 54 AA;

Query Match	53.2%	Score 42;	DB 5;	Length 54;
Best Local Similarity	58.3%	Score No. 11;		
Matches	7;	Conservative	2;	Mismatches 3; Indels 0; Gaps 0;

```

QY      1 HITPAGSSNYY 12
          | : | | | | : |
Db      39 HLHPXGSSNLY 50

```

RESULT 10  
ABB64072  
ID ABB64072 standard; protein; 797 AA

AC ABB64072;

DT 26-MAR-2002 (first entry)

Drosophila melanogaster SEQ ID NO 19008

kw Drosophila; developmental biology; cell signalling; insecticide resistance

XX  
XX  
Dyodanrhij a mej anogastay

XX WO200171042-A2  
DN

XX PD 27-SEP-2001.

23-MAR-2001: 2001WO-US009231.

23-MAR-2000; 2000US-0191637P.

7	H
8	H
9	C
10	C
11	E
12	M
13	C
14	C
15	"
16	M
17	C
18	C
19	C
20	C
21	C
22	<
23	C
24	C
25	F
26	M
27	F
28	U

X  
X  
X

(  
E  
B  
C  
S  
I

XX

DR N-PSDB; ABL08175.

PT	New isolated nucl

PT interactions.

PS Disclosure; SEQ ID NO 19008; 21pp + Sequence Listing; English.

The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from *Drosophila*. The invention is



RESULT 13  
 OS ABB64461  
 ID ABB64461 standard; protein; 3508 AA.  
 XX  
 AC ABB64461;  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE Drosophila melanogaster polypeptide SEQ ID NO 20175.  
 XX  
 KW Drosophila; developmental biology; cell signalling; insecticide;  
 XX pharmaceutical.  
 XX  
 PA Drosophila melanogaster.  
 OS  
 PN WO200171042-A2.  
 XX  
 PD 27-SEP-2001.  
 XX  
 PF 23-MAR-2001; 2001WO-US009231.  
 XX  
 PR 23-MAR-2000; 2000US-0191637P.  
 XX 11-JUL-2000; 2000US-00614150.  
 XX  
 PA (PEKE ) PE CORP NY.  
 XX  
 PI Venter JC, Adams M, Li PMD, Myers EW;  
 XX WPI; 2001-656860/75.  
 DR N-PSDB; ABL08564.  
 XX  
 PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signaling and cell-cell  
 PT interactions.  
 XX  
 PS Disclosure; SEQ ID NO 20175; 21pp + Sequence Listing; English.  
 XX  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (AB116176-AB10511), expressed DNA  
 CC sequences (AB101840-AB116175) and the encoded proteins (AB57737-  
 CC AB572072). The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 3508 AA;  
 Query Match 51.9%; Score 41; DB 4; Length 3508;  
 Best Local Similarity 42.9%; Pred. No. 2e+03;  
 Matches 6; Conservative 3; Mismatches 5; Indels 0; Gaps 0;  
 QY 1 HITPAGSSNYVY 14  
 DB 3159 HVTPKSYLNFAGY 3172

RESULT 14  
 OS AAO06759  
 ID AAO06759 standard; protein; 61 AA.  
 XX  
 AC AAO06759;  
 XX  
 DT 06-NOV-2001 (first entry)  
 XX  
 DE Human polypeptide SEQ ID NO 20651.  
 XX  
 KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
 KW nervous system disorders; arthritis; inflammation.

XX Homo sapiens.  
 OS  
 XX WO200164835-A2.  
 XX  
 PN  
 XX  
 PD 07-SEP-2001.  
 XX  
 PF 26-FEB-2001; 2001WO-US004927.  
 XX  
 PR 28-FEB-2000; 2000US-00515126.  
 XX 18-MAY-2000; 2000US-00577409.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 OS  
 PN Tang YT, Liu C, Drmanac RT;  
 XX WPI; 2001-514838/56.  
 DR N-PSDB; AA186690.  
 XX  
 PT Isolated nucleic acids and polypeptides, useful for preventing diagnosing  
 PT and treating e.g. leukemia, inflammation and immune disorders.  
 XX  
 PS Claim 20; SEQ ID NO 20651; 1399pp + Sequence Listing; English.  
 XX  
 CC The invention relates to human polynucleotides (AA179941-AA193841) and  
 CC the encoded proteins (AA000010-AA013910) that exhibit activity relating to  
 CC cytokine, cell proliferation or cell differentiation or which may induce  
 CC production of other cytokines in other cell populations. The  
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
 CC peptide therapy. The polypeptides have various cytokine-like activities,  
 CC e.g. stem cell growth factor activity, haematopoiesis regulating  
 CC activity, tissue growth factor activity, immunomodulatory activity and  
 CC activity/inhibin activity and may be useful in the diagnosis and/or  
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
 CC inflammation. Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic format  
 CC directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 61 AA;  
 Query Match 50.6%; Score 40; DB 4; Length 61;  
 Best Local Similarity 66.7%; Pred. No. 28;  
 Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
 QY 1 HITPAGSSNYVY 12  
 DB 13 HQTPALSSNFFY 24

RESULT 15  
 OS ADE09113  
 ID ADE09113 standard; protein; 242 AA.  
 XX  
 AC ADE09113;  
 XX  
 DT 29-JAN-2004 (first entry)  
 XX  
 DE Novel protein-related contig polypeptide sequence #1179.  
 XX  
 KW novel gene; novel protein; tissue marker; molecular weight marker;  
 KW chromosome marker; genetic disorder; contig.  
 XX  
 OS Unidentified.  
 XX  
 PN WO2003054152-A2.  
 XX  
 PD 03-JUL-2003.  
 XX  
 PF 10-DEC-2002; 2002WO-US039555.  
 XX  
 PR 10-DEC-2001; 2001US-0339739P.  
 XX 11-DEC-2001; 2001US-0339453P.  
 PR 14-MAR-2002; 2002US-0365091P.

PR 14-MAR-2002; 2002US-0365384P.  
 PR 12-APR-2002; 2002US-0372381P.  
 PR 12-APR-2002; 2002US-0372615P.  
 PR 22-APR-2002; 2002US-00128558.  
 PR 24-APR-2002; 2002US-0376045P.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Asundi V, Goodrich RW, Ren F, Zhang J, Zhao Q, Wang J;  
 PI Ghosh M, Xue AJ, Wehrman T, Meng G, Zhou P, Drmanac RT, Wang Z;  
 PI Ma Y, Wang D, Chen R, Xu C, Boyle BJ;  
 XX

DR WPI; 2003-569235/53.

XX New polynucleotides, useful for expressing recombinant proteins for  
 PT analysis, characterization or therapeutic use, or as markers for tissues  
 PT in which the corresponding protein is preferentially expressed.

PS Disclosure; SEQ ID NO 2657; 1177bp; English.

XX The invention comprises the amino acid and coding sequences of novel  
 CC proteins. The DNA and protein sequences of the invention are useful as:  
 CC markers for tissues in which the corresponding protein is preferentially  
 CC expressed; as molecular weight markers on gels; as chromosome markers or  
 CC tags; to identify chromosomes or to map related gene positions; and to  
 CC compare with endogenous DNA sequences in patients to identify potential  
 CC genetic disorders. The present amino acid sequence was used in the  
 CC exemplification of the invention.

XX Sequence 242 AA;

Query Match

50.6%; Score 40; DB 7; Length 242;

Best Local Similarity 72.7%; Pred. No. 1.4e+02; Mismatches 0; Indels 0;

OY 4 PAGSSNVYGY 14  
 |||||  
 Db 38 PAGSSGYRAGY 48

Search completed: October 5, 2004, 08:00:03  
 Job time : 14.6512 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2004, 07:36:16 ; Search time 2.10648 Seconds  
(without alignments)  
593.639 Million cell updates/sec

Title: US-09-805-290A-17  
Perfect score: 72  
Sequence: 1 DIRKRTSGVSHY 13

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 283366 seqs, 96191526 residues  
Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum March 0%  
Maximum March 100%  
Listing first 45 summaries

Database :  
1: PIR.78:\*  
2: PIR1:\*  
3: PIR2:\*  
4: PIR3:\*  
5: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	43	59.7	759	2	D70422 cellulose synthase
2	43	59.7	780	2	T16679 hypothetical prote
3	42	58.3	401	2	D81366 probable integral
4	40	55.6	218	2	T30415 host range factor
5	40	55.6	250	2	AP2881 lipase esterase [i
6	40	55.6	250	2	F97657 lipase esterase (A
7	40	55.6	459	2	T24425 cytochrome b homol
8	40	55.6	549	2	E70784 cytochrome b homol
9	40	55.6	551	2	A87019 probable cytochrom
10	40	55.6	732	2	A64107 glycosyltransferas
11	39.5	54.2	472	2	T30101 hypothetical prote
12	39	54.2	195	2	T24033 hypothetical prote
13	39	54.2	205	2	H70468 phosphoribosyl-AMP
14	39	54.2	258	1	WMBPP9 lysozyme (EC 3.2.1
15	39	54.2	258	1	WMBP15 lysozyme (EC 3.2.1
16	39	54.2	337	2	B95423 RepB2 replication
17	39	54.2	431	2	A89761 hypothetical prote
18	39	54.2	440	2	T44138 hypothetical prote
19	39	54.2	605	2	H86640 protein f52C12.2 [
20	39	54.2	758	2	H75013 hypothetical prote
21	39	54.2	825	2	H64083 biotin sulfoxide r
22	39	54.2	1127	2	T30334 AND-1 protein - Af
23	39	54.2	1253	2	F86436 hypothetical prote
24	38.5	53.5	284	2	F97216 probable aminoglyc
25	38	52.8	300	2	A82932 ABC Transporter UU
26	38	52.8	349	2	C59107 hypothetical prote
27	38	52.8	478	2	S44249 bet protein - huma
28	38	52.8	507	2	T08337 hypothetical prote
29	38	52.8	683	2	D69167 DNA helicase II -

30	38	52.8	880	2	F83386 hypothetical prote
31	38	52.8	1371	2	T29019 hypothetical prote
32	37	51.4	144	2	A69859 conserved hypothet
33	37	51.4	157	2	T24380 hypothetical prote
34	37	51.4	204	2	T18865 hypothetical prote
35	37	51.4	207	2	AC3381 hypothetical prote
36	37	51.4	265	2	US0260 serine proteinase
37	37	51.4	265	2	A38894 kinase-related tra
38	37	51.4	283	1	TVVZ82 kinase-related tra
39	37	51.4	283	1	T37448 protein kinase - v
40	37	51.4	283	2	T37448 hypothetical prote
41	37	51.4	469	2	S76891 conserved hypothet
42	37	51.4	471	2	A81680 conserved hypothet
43	37	51.4	502	2	T25263 hypothetical prote
44	37	51.4	528	2	B95339 FikI-related hist
45	37	51.4	533	2	T33182 hypothetical prote

## ALIGNMENTS

RESULT 1  
D70422 cellulose synthase catalytic subunit - Aquifex aeolicus  
C/Species: Aquifex aeolicus  
C/Date: 08-May-1998 #sequence\_revision 08-May-1998 #text\_change 05-Nov-1999  
C/Accession: D70422  
R/Decker, G., Warren, P.V., Gaasterland, T., Young, W.G., Lenox, A.L., Graham, D.E., Ove  
V.  
Nature 392, 353-358, 1998  
A/Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.  
A/Reference number: A70300; PMID:9519666; PMID:9537320  
A/Accession: D70422  
A/Status: preliminary; nucleic acid sequence not shown; translation not shown  
A/Molecule type: DNA  
A/Residues: 1-759 <NO>  
A/Cross-references: GB:AE000738; NID:g2983801; PIDN:AA007360.1; PID:g2983803; GB:AE000657  
A/Experimental source: strain Vfs  
C/Genetics:  
A/Gene: bcsA  
C/Superfamily: bcsA protein

Query Match 59.7%; Score 43; DB 2; Length 759;  
Best Local Similarity 70.0%; Pred. No. 12;  
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 4 KRFTSGVSHY 13  
DB 65 KRFTSGVSHY 74

## RESULT 2

T16679 hypothetical protein R04A9.5 - Caenorhabditis elegans  
C/Species: Caenorhabditis elegans  
C/Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 03-Jun-2002  
C/Accession: T16679  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Accession: T16679  
A/Molecule type: DNA  
A/Residues: 1-780 <GB>  
A/Cross-references: EMBL:U01550; NID:g1118045; PID:g1118049; PIDN:AA83287.1; CESP:R04A9.  
C/Genetics:  
A/Gene: CESP:R04A9.5  
A/Intons: 36/3; 326/3; 349/3; 377/3; 396/3; 428/3; 484/3; 515/1; 632/3; 671/1; 752/3

Query Match 59.7%; Score 43; DB 2; Length 780;  
Best Local Similarity 80.0%; Pred. No. 12;  
Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 RKRFTSGYSH 12  
 |||||  
 Db 622 RKRFTGLGYH 631

## RESULT 3

probable integral membrane protein Cj0941c [imported] - Campylobacter jejuni (strain NCTC 13627)  
 C/Species: Campylobacter jejuni  
 C/Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 03-Jun-2002  
 C/Accession: D81368  
 R/Parikh, J.; Wren, B.W.; Mungall, K.; Kellley, J.M.; Churcher, C.; Basham, D.; Chittam  
 C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; Varghese, A.; Whitehead, S.; Barral  
 Nature 403, 665-668, 2000  
 A/Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hyp  
 A/Reference number: A81250; WUID:20150912; PMID:10688204  
 A/Accession: D81368  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-401 <PAR>  
 A/Cross-references: GB:AL139076; GB:AL111168; NID:g6868128; PIDN:CAB73198.1; PID:g686837  
 A/Experimental source: serotype O2, strain NCTC 11168  
 C/Genetics:  
 A/Gene: Cj0941c  
 C/Superfamily: hypothetical protein H1155

Query Match 58.3%; Score 42; DB 2; Length 401;  
 Best Local Similarity 53.8%; Pred. No. 9.2;  
 Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;  
 QY 1 DIRKRTSGYSH 13  
 |||||  
 Db 183 DVKARTSLGYH 195

## RESULT 4

host range factor 1 - Lymantia dispar nuclear polyhedrosis virus  
 C/Species: Lymantia dispar nuclear polyhedrosis virus, LDMNPV  
 C/Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 21-Jul-2000  
 C/Accession: T10415  
 R/Kuzio, J.; Pearson, M.N.; Harwood, S.H.; Funk, C.J.; Evans, J.T.; Slaveck, J.M.; Roh  
 Virology 253, 17-34, 1999  
 A/Title: Sequence and analysis of the genome of a baculovirus pathogenic for Lymantia d  
 A/Reference number: Z20836; WUID:99124785; PMID:9887315  
 A/Accession: T10415  
 A/Status: preliminary; translated from GB/EMBL/DBJ  
 A/Molecule type: DNA  
 A/Residues: 1-218 <KUZ>  
 A/Cross-references: EMBL:AF081810; NID:g3822234; PIDN:AACT0253.1; PID:g3822302  
 C/Superfamily: Lymantia dispar nuclear polyhedrosis virus host range factor 1

Query Match 55.6%; Score 40; DB 2; Length 218;  
 Best Local Similarity 50.0%; Pred. No. 11;  
 Matches 6; Conservative 2; Mismatches 4; Indels 0; Gaps 0;  
 QY 2 IRKRTSGYSH 13  
 |||||  
 Db 37 VRYRERGGYH 48

## RESULT 5

lipase esterase [imported] - Agrobacterium tumefaciens (strain C58, Dupont)  
 C/Species: Agrobacterium tumefaciens  
 C/Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 18-Nov-2002  
 C/Accession: AF2881  
 R/Wood, D.M.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woc, I  
 erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; Mccl  
 erage, P.; Romero, J.; Zhang, S.  
 Science 294, 2317-2323, 2001  
 A/Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,

ster, E.W.  
 A/Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.  
 A/Reference number: AB2577; WUID:21608550; PMID:11743193  
 A/Accession: AF2881  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-250 <KUR>  
 A/Cross-references: GB:AE008688; PIDN:AAL43468.1; PID:g17740973; GSPDB:GN00186  
 A/Experimental source: strain C58 (Dupont)  
 C/Genetics:  
 A/Gene: Atu2481  
 A/Map position: circular chromosome

Query Match 55.6%; Score 40; DB 2; Length 250;  
 Best Local Similarity 50.0%; Pred. No. 13;  
 Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;  
 QY 2 IRKRTSGYSH 13  
 |||||  
 Db 146 MRQRMADGMSHY 157

## RESULT 6

lipase esterase (AP001514) [imported] - Agrobacterium tumefaciens (strain C58, Cereon)  
 C/Species: Agrobacterium tumefaciens  
 C/Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 18-Nov-2002  
 C/Accession: F97657  
 R/Goodier, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quirillo, B.; Goldman,  
 A.; Liu, F.; Wollam, C.; Allinger, M.; Doughy, D.; Scott, C.; Lappas, C.; Mackeltz, B.;  
 Science 294, 2323-2328, 2001  
 A/Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tume  
 A/Reference number: A97359; WUID:21608551; PMID:11743194  
 A/Accession: F97657  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-250 <KUR>  
 A/Cross-references: GB:AE007869; PIDN:AAK88215.1; PID:g15157667; GSPDB:GN00169  
 C/Genetics:  
 A/Gene: AGR C 4503  
 A/Map position: circular chromosome

Query Match 55.6%; Score 40; DB 2; Length 250;  
 Best Local Similarity 50.0%; Pred. No. 13;  
 Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;  
 QY 2 IRKRTSGYSH 13  
 |||||  
 Db 146 MRQRMADGMSHY 157

## RESULT 7

hypothetical protein T04A8.3 - Caenorhabditis elegans  
 C/Species: Caenorhabditis elegans  
 C/Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
 C/Accession: T24425  
 R/Palmer, S.  
 submitted to the EMBL Data Library, August 1994  
 A/Reference number: Z19889  
 A/Accession: T24425  
 A/Status: preliminary; translated from GB/EMBL/DBJ  
 A/Molecule type: DNA  
 A/Residues: 1-459 <WIL>  
 A/Cross-references: EMBL:Z35663; PIDN:CAA84726.1; GSPDB:GN00021; CESP:T04A8.3  
 A/Experimental source: clone T04A8  
 C/Genetics:  
 A/Gene: CESP.T04A8.3  
 A/Map position: 3  
 A/Introns: 25/2; 43/2; 130/1; 272/1; 326/1

Query Match 55.6%; Score 40; DB 2; Length 459;  
 Best Local Similarity 63.6%; Pred. No. 25;

```
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
QY 1 DIRKRTSGYS 11
|:|:|:|:|
Db 352 DLRKFTGYS 362

RESULT 8
cytochrome b homolog cytb [similarity] - Mycobacterium tuberculosis (strain H37RV)
E70784
C/Species: Mycobacterium tuberculosis
C/Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Apr-2000
C/Accession: E70784
R/Col: S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A/Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A/Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A/Reference number: A70500; PMID:96255987; PMID:9634230
A/Accession: E70784
A/Status: preliminary; nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-549 <COL>
A/Cross-references: GB:270283; GB:AL123456; NID:g3261561; PIDN:CAA94265.1; PID:e233572;
A/Experimental source: strain H37RV
C/Genetics:
A/Gene: qcrB
C/Keywords: heme; iron; metalloprotein (axial ligands) (low potential) #status predicted
F/114,216/Binding site: heme iron (His) (axial ligands) (high potential) #status predicted
F/128,231/Binding site: heme iron (His)
Query Match 55.6%; Score 40; DB 2; Length 549;
Best Local Similarity 50.0%; Pred. No. 29;
Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 2 IRKRFTSGYSHY 13
|:|:|:|:|
Db 355 LERKFTGDYAH 366

RESULT 9
probable cytochrome b [imported] - Mycobacterium leprae
A87019
C/Species: Mycobacterium leprae
C/Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C/Accession: A87019
R/Col: S.T.; Eigmeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; Ho R.; Davies, R.M.; Devlin, K.; Duthey, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holroyd, S.; Cole, M.A.; Rutherford, K.M.
Nature 409, 1007-1011, 2001
A/Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; Sq A/Title: Massive gene decay in the leprosy bacillus.
A/Reference number: A86909; PMID:21128732; PMID:11234002
A/Accession: A87019
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-551 <STO>
A/Cross-references: GB:AL450380; NID:g13092952; PIDN:CAC31260.1; GSPDB:GN00147
C/Genetics:
A/Gene: qcrB

Query Match 55.6%; Score 40; DB 2; Length 551;
Best Local Similarity 50.0%; Pred. No. 30;
Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 2 IRKRFTSGYSHY 13
|:|:|:|:|
Db 354 LERKFTGDYAH 365

RESULT 10
A84107
```

```
glycosyltransferase BH3657 [imported] - Bacillus halodurans (strain C-125)
C/Species: Bacillus halodurans
C/Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C/Accession: A84107
R/Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fuji, F.; Hirai Nucleic Acids Res. 28, 4317-4331, 2000
A/Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and A/Reference number: A83650; PMID:20512582; PMID:11058132
A/Accession: A84107
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-732 <STO>
A/Cross-references: GB:AP001519; GB:BA000004; NID:g10176109; PIDN:BA07376.1; GSPDB:GN001 A/Experimental source: strain C-125
C/Genetics:
A/Gene: BH3657

Query Match 55.6%; Score 40; DB 2; Length 732;
Best Local Similarity 50.0%; Pred. No. 40;
Matches 6; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 2 IRKRFTSGYSHY 13
|:|:|:|:|
Db 326 VKRKYFDSYSHY 337

RESULT 11
hypothetical protein C06G3.1 - Caenorhabditis elegans
T30101
C/Species: Caenorhabditis elegans
C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C/Accession: T30101
R/Latreille, P.; Scelley, L.
submitted to the EMBL Data Library, June 1996
A/Description: The sequence of C. elegans cosmid C06G3.
A/Reference number: Z20735
A/Accession: T30101
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-472 <LAT>
A/Cross-references: EMBL:U61947; PIDN:AAB03131.1; GSPDB:GN00022; CESP:C06G3.1
C/Genetics:
A/Gene: CESP:C06G3.1
A/Map position: 4
A/Intons: 21/1; 127/1; 234/3; 276/3; 306/2; 375/1; 427/3

Query Match 54.9%; Score 39.5; DB 2; Length 472;
Best Local Similarity 57.1%; Pred. No. 31;
Matches 8; Conservative 2; Mismatches 3; Indels 1; Gaps 1;

QY 1 DIRKRTSGYSHY 13
|:|:|:|:|
Db 398 DVNRVINGLYSHY 411

RESULT 12
T24033
hypothetical protein R07E3.7 - Caenorhabditis elegans
C/Species: Caenorhabditis elegans
C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 16-Feb-2000
C/Accession: T24033
R/Cottage, A.
submitted to the EMBL Data Library, May 1995
A/Reference number: Z19832
A/Accession: T24033
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-195 <WIL>
A/Cross-references: EMBL:Z49207; PIDN:CAA89073.1; GSPDB:GN00028; CESP:R07E3.7
A/Experimental source: clone R07E3
C/Genetics:
A/Gene: CESP:R07E3.7
```

A:Map position: X  
A:introns: 46/1; 81/1; 113/3; 147/1

Query Match 54.2%; Score 39; DB 2; Length 195;  
Best Local Similarity 50.0%; Pred. No. 16;  
Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 DIRKRTSGYSH 12  
|:|:|:|:|:  
DB 86 DSKRSTAFSH 97

## RESULT 13

H70468 phosphoribosyl-AMP cyclohydrolase (EC 3.5.4.19) / phosphoribosyl-ATP diphosphatase (EC 3

C/Species: Aquifex aeolicus

C/Date: 08-May-1998 #sequence\_revision 08-May-1998 #text\_change 16-Aug-2002

C/Accession: H70468

R/Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; O'V

Nature 392, 353-358, 1998

A/Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.

A/Reference number: A70300; MUID:98196666; PMID:9537320

A/Accession: H70468

A/Status: preliminary; nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-205 <AOP>

A/Cross-references: GB:AEO00765; NID:G2984193; PIDN:AAC07730.1; PID:G2984200; GB:AEO0065

A/Experimental source: strain VFS

C/Genetics:

A/Gene: hsiB

C/Superfamily: hsiB bifunctional enzyme; hsiB bifunctional enzyme homology; hsiB protein

C/Keywords: hydrolase

F/2-205/Domain: hsiB bifunctional enzyme homology <HSI>

F/118-189/Domain: hsiB protein homology <HSI>

Query Match 54.2%; Score 39; DB 2; Length 205;

Best Local Similarity 50.0%; Pred. No. 16;

Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 2 IRRKRTSGYSH 13  
|:|:|:|:|:  
DB 35 IKRTLETGYAHY 46

## RESULT 14

KMBPP9

Lysozyme (EC 3.2.1.17) - phage phi-29

N/Alternate names: morphogenesis protein 2

C/Species: phage phi-29

A/Note: host Bacillus amyloliquefaciens; Bacillus subtilis

C/Date: 31-Mar-1989 #sequence\_revision 31-Mar-1989 #text\_change 18-Jun-1999

C/Accession: B24721; A33078

R/Garvey, K.J.; Saeed, M.S.; Ito, J.

Nucleic Acids Res. 14, 10001-10008, 1986

A/Title: Nucleotide sequence of Bacillus phage phi-29 genes 14 and 15: homology of gene

A/Reference number: A24721; MUID:87117505; PMID:3027653

A/Accession: B24721

A/Molecule type: DNA

A/Residues: 1-258 <GAR>

A/Cross-references: GB:X04962; NID:G15676; PIDN:CAA28632.1; PID:G15679

R/Vick, C.; Paces, V.

Gene 46, 215-225, 1986

A/Title: Nucleotide sequence of the late region of Bacillus phage phi-29 completes the 1

A/Reference number: A25816; MUID:87106857; PMID:3803926

A/Accession: A33078

A/Molecule type: DNA

A/Residues: 1-258 <VLC>

A/Cross-references: GB:M44782; NID:G215323; PIDN:AAA32288.1; PID:G215333

C/Genetics:

A/Gene: 15

C/Function:

A/Description: catalyzes hydrolysis of the beta-1,4-glycosidic bond between N-acetylmutan  
C/Superfamily: phage T4 lysozyme; phage T4 lysozyme homology  
C/Keywords: glycosidase; host cell lysis; lysozyme; late protein  
F/15-134/Domain: phage T4 lysozyme homology <T4L>

Query Match 54.2%; Score 39; DB 1; Length 258;  
Best Local Similarity 60.0%; Pred. No. 21;  
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 4 KRFTSGYSHY 13  
|:|:|:|:|:  
DB 29 KHYTIGYGHY 38

## RESULT 15

WMBP15

Lysozyme (EC 3.2.1.17) - phage PZA

C/Species: phage PZA

A/Note: host Bacillus subtilis

C/Date: 31-Mar-1988 #sequence\_revision 31-Mar-1988 #text\_change 18-Jun-1999

C/Accession: A26215; J24831

R/Paces, V.; Vick, C.; Urbanek, P.

Gene 44, 107-114, 1986

A/Title: Nucleotide sequence of the late region of Bacillus subtilis phage PZA, a close ;

A/Reference number: A91550; MUID:87031573; PMID:3095188

A/Accession: A26215

A/Molecule type: DNA

A/Residues: 1-258 <PAC>

A/Cross-references: GB:M1813; GB:M13904; GB:M13905; NID:G216046; PIDN:AAA88492.1; PID:G;

C/Genetics:

A/Gene: 15

C/Function:

A/Description: catalyzes hydrolysis of the beta-1,4-glycosidic bond between N-acetylmutan

C/Superfamily: phage T4 lysozyme; phage T4 lysozyme homology

C/Keywords: glycosidase; host cell lysis; hydrolase; late protein

F/15-134/Domain: phage T4 lysozyme homology <T4L>

Query Match 54.2%; Score 39; DB 1; Length 258;  
Best Local Similarity 60.0%; Pred. No. 21;  
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 4 KRFTSGYSHY 13  
|:|:|:|:|:  
DB 29 KHYTIGYGHY 38

Search completed: October 5, 2004, 08:17:12  
Job time : 5.10648 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2004, 07:29:35 ; Search time 1.16358 Seconds  
(without alignments)  
581.749 Million cell updates/sec

Title: US-09-805-290A-17  
Perfect score: 72  
Sequence: 1 DIRKRTSGYSHY 13

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues  
Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Swisssprot\_42: \*  
Pred. NO. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

Result No.	Score	Query Match	Length	ID	Description
1	40	55.6	218	HRF1_NPVLD	Q90165 lymphatic d
2	40	55.6	549	OCRB_MYCTU	Q10388 mycobacteri
3	40	55.6	551	OCRB_MYCTU	P15878 mycobacteri
4	40	55.6	970	AND1_MOUSE	P59328 mus musculu
5	39	54.2	205	HIS2_MOUSE	O67780 aquifex aeo
6	39	54.2	258	LYCV_BPPI2	P11187 bacterioph
7	39	54.2	258	LYCV_BPPI2	P07540 bacterioph
8	39	54.2	341	ARGC_STAP2	Q63535 staphylococ
9	39	54.2	825	TORZ_HAELN	P44798 haemophilus
10	39	54.2	1127	AND1_XENLA	Q13046 xenopus lae
11	37	51.4	141	HIS3_METKA	P58836 methanopyru
12	37	51.4	263	LYCV_BPPI3	Q37886 bacterioph
13	37	51.4	263	SER1_DROME	P17205 diosophila
14	37	51.4	283	KRB2_VACCV	P21098 vaccinia vi
15	37	51.4	283	KRB2_VACCV	P24362 vaccinia vi
16	37	51.4	812	AXN2_BRAAE	P57095 brachydanio
17	37	51.4	1129	AND1_HUMAN	O75717 homo sapien
18	37	51.4	1222	SOR3_HUMAN	Q93943 homo sapien
19	36.5	50.7	285	VAPY_ECOLI	P77325 escherichia
20	36	50.0	95	YGHM_ECOLI	Q46848 escherichia
21	36	50.0	215	LFTF_CAMOE	Q90165 mycobacteri
22	36	50.0	241	KDNM_BPT4	Q90165 mycobacteri
23	36	50.0	507	TRA6_PSEAE	O57541 pseudomonas
24	36	50.0	529	VSM6_TRYBB	P26334 trypanosoma
25	36	50.0	562	MTRB_MYCIE	Q90165 mycobacteri
26	36	50.0	565	MTRB_MYCIE	Q90165 mycobacteri
27	36	50.0	567	MTRB_MYCBO	P38963 mycobacteri
28	36	50.0	567	MTRB_MYCTU	Q50496 mycobacteri
29	36	50.0	729	COA1_PAVP9	P23448 porcine par
30	36	50.0	729	COA1_PAVP9	P55010 porcine par
31	36	50.0	1115	PAW2_YEAST	P55010 ascaromyc
32	35	48.6	127	HIS3_METKA	O58825 methanococ
33	35	48.6	240	HXA1_XENLA	Q08821 xenopus lae

34	35	48.6	253	1	PRPK_HUMAN	Q96544 homo sapien
35	35	48.6	323	1	PRIS_METTH	O26685 methanobact
36	35	48.6	326	1	WRK8_ARATH	O51126 arabidopsis
37	35	48.6	471	1	YHU6_YEAST	P18770 saccharomyc
38	35	48.6	502	1	RXU1_SEOCE	O03723 secale cere
39	35	48.6	558	1	Y322_MYCCE	P47564 mycoplasma
40	35	48.6	576	1	ORC2_MOUSE	Q60862 mus musculu
41	35	48.6	667	1	M110_CAEBL	P34400 caenorhabdi
42	35	48.6	955	1	FRU_DROME	Q81n81 diosophila
43	35	48.6	1015	1	TNP3_ECOLI	Q03008 escherichia
44	34.5	47.9	786	1	AAC_ACTUT	P23958 actinoplan
45	34	47.2	132	1	YBH8_PSEAE	Q9hv30 pseudomonas

ALIGNMENTS

RESULT 1  
ID HRF1\_NPVLD STANDARD; PRT; 218 AA.  
AC Q90165;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 15-JUL-1999 (Rel. 36, Last annotation update)  
DE Host range factor 1.  
GN HRF-1.  
OS Lymantia dispar multicapsid nuclear polyhedrosis virus (LdMNPV).  
OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae;  
OC Nucleopolyhedrovirus.  
OX NCBI\_TaxID=10449;  
RN [1]  
RC STRAIN=Hamden; RP  
RX MEDLINE=98139158; PubMed=9499118;  
RA Chen C.J., Quentin M.E., Brennan L.A., Kukul C., Thiem S.M.;  
RT "Lymantia dispar nucleopolyhedrovirus hrf-1 expands the larval host  
range of Autographa californica nucleopolyhedrovirus";  
RL J. Virol. 72:2526-2531 (1998).  
CC -!- FUNCTION: Facilitates AcMNPV replication in two non-permissive  
cell lines, IPLB-Ld652Y and IPLB-LdFB.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; U38895; AA07701.1; -  
DR EMBL; U38895; AA07701.1; -  
DR EMBL; U38895; AA07701.1; -  
SQ SEQUENCE 218 AA; 25675 MW; 5DD2B80E1602001 CRC64;  
  
Query Match 55.6%; Score 40; DB 1; Length 218;  
Best Local Similarity 50.0%; Pred. No. 4.7;  
Matches 6; Conservative 2; Mismatches 4; Indels 0; Gaps 0;  
  
Db 2 IRKRRTSGYSHY 13  
37 VRYRYERGGYHY 48  
  
RESULT 2  
ID QCRB\_MYCTU STANDARD; PRT; 549 AA.  
AC Q10388;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Ubiqutinol-cytochrome c reductase cytochrome B subunit.  
GN QCRB OR RV2196 OR MT2252 OR MTCY190.07 OR MB2219.  
OS Mycobacterium tuberculosis, and  
OS Mycobacterium bovis.



CC or send an email to license@isb-sib.ch).  
CC -----  
CC EMBL; AL583920; CAC31260.1; -  
DR EMBL; X51328; CAA35710.1; ALT\_FRAME.  
DR PIR; A87019; A87019.  
DR Lepidoma; M10879; -  
DR InterPro; IPR005797; Cytochrome b6 N.  
DR Pfam; PF00033; Cytochrome b6 N; 1.  
DR POSITE; PS00192; CYTOCHROME B HEME; FALSE NEG.  
KM Electron transport; Heme; Transmembrane; Complete proteome.  
FT TRANSMEM 44 64 POTENTIAL.  
FT TRANSMEM 117 137 POTENTIAL.  
FT TRANSMEM 145 165 POTENTIAL.  
FT TRANSMEM 188 208 POTENTIAL.  
FT TRANSMEM 216 236 POTENTIAL.  
FT TRANSMEM 265 285 POTENTIAL.  
FT TRANSMEM 334 354 POTENTIAL.  
FT TRANSMEM 380 400 POTENTIAL.  
FT TRANSMEM 417 437 POTENTIAL.  
FT METAL 113 113 IROM (HEME 1 AXIAL LIGAND) (BY SIMILARITY).  
FT METAL 127 127 IROM (HEME 2 AXIAL LIGAND) (BY SIMILARITY).  
FT METAL 215 215 IROM (HEME 1 AXIAL LIGAND) (BY SIMILARITY).  
FT METAL 230 230 IROM (HEME 2 AXIAL LIGAND) (BY SIMILARITY).  
SQ SEQUENCE 551 AA; 61545 MW; 65BDC841D1DD3239 CRC64;  
Query Match 55.6%; Score 40; DB 1; Length 551;  
Best Local Similarity 50.0%; Pred. No. 13;  
Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;  
QY 2 IRKRFITSGYSHY 13  
Db 354 LEKRFITGDYAH 365  
RESULT 4  
AND1\_MOUSE STANDARD; PRT; 970 AA.  
AC PS9328;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Acidic nucleoplasmic DNA-binding protein 1 (And-1) (Fragment).  
GN AND1.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Bone, and Kidney;  
RX MEDLINE=21085660; PUBMED=11217851;  
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
RA Aikawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,  
RA Aikawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana I.,  
RA Saito T., Okazaki Y., Gotojori T., Bono H., Kasukawa T., Saito R.,  
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochwa H.,  
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,  
RA Schiraldi L.M., Staib F., Suzuki R., Tomita K., Wagner L., Washio T.,  
RA Sakai K., Okido T., Futuno M., Aono H., Baldarelli R., Barsh G.,  
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
RA Gustincich S., Hill D., Hofmann C., Fujita M., Gariboldi M.,  
RA Lyons P., Marchionni L., Washima J., Marzelli U., Nombarelli P.,  
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
RA Saeki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,  
RA Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,  
RA Hayashizaki Y.,

RT "Functional annotation of a full-length mouse cDNA collection."  
RL Nature 409:685-690(2001).  
CC -1- FUNCTION: Binds DNA with high affinity. May also be involved in  
CC protein-protein interactions (By similarity).  
CC -1- SUBCELLULAR LOCATION: Nuclear; nucleoplasm (By similarity).  
CC -1- ALTERNATIVE PRODUCTS:  
CC Zevent-Alternative splicing; Named isoforms=2;  
CC Name=1;  
CC IsoId=P59328-1; Sequence=Displayed;  
CC Name=2;  
CC IsoId=P59328-2; Sequence=VSP 006757, VSP 006758;  
CC Note=No experimental confirmation available;  
CC -1- SIMILARITY: Contains 7 WD repeats.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to license@isb-sib.ch).  
CC -----  
CC EMBL; AK036390; BAC29408.1; -  
CC EMBL; AK052690; BAC35097.1; -  
DR MGD; MGI:2443514; D630024B06R1k.  
DR InterPro; IPR001660; WD40.  
DR Pfam; PF00400; WD40; 4.  
DR ProDom; PD000018; WD40; 1.  
DR PROSITE; PS00678; WD\_REPEATS\_1; 2.  
DR PROSITE; PS00678; WD\_REPEATS\_2; 3.  
DR PROSITE; PS02994; WD\_REPEATS\_REGION; 1.  
KW Nuclear protein; DNA-binding; Repeat; WD repeat; Alternative splicing.  
FT REPEAT 11 50 WD 1.  
FT REPEAT 52 91 WD 2.  
FT REPEAT 93 131 WD 3.  
FT REPEAT 134 173 WD 4.  
FT REPEAT 184 223 WD 5.  
FT REPEAT 228 267 WD 6.  
FT REPEAT 271 310 WD 7.  
FT VARSPLIC 630 641 TPCYVDSEGCVR -> KIIFFLYINDIF (in isoform 2)  
FT FT VARSPLIC 642 >970 /FTId=VSP 006757.  
FT FT VARSPLIC 642 >970 Missing (in isoform 2).  
FT FT CONFLICT 54 54 /FTId=VSP 006758.  
FT FT CONFLICT 350 350 G -> V (IN REF. 1; BAC29408).  
FT NON\_TER 970 970 S -> P (IN REF. 1; BAC29408).  
SQ SEQUENCE 970 AA; 107632 MW; 2B2B577A9A6D2792 CRC64;  
Query Match 55.6%; Score 40; DB 1; Length 970;  
Best Local Similarity 50.0%; Pred. No. 23;  
Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;  
QY 1 DIRKRTSGYSH 12  
Db 828 DFEKKNAGYSH 839  
RESULT 5  
HIS2\_AQUAE STANDARD; PRT; 205 AA.  
ID HIS2\_AQUAE  
AC 067780;  
DT 15-DEC-1998 (Rel. 37, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 28-FEB-1998 (Rel. 41, Last annotation update)  
DE Histidine biosynthesis bifunctional protein his2E (includes:  
DE Phosphoribosyl-AMP cyclohydrolase (EC 3.5.4.19) (PRA-CH);  
DE Phosphoribosyl-ATP pyrophosphatase (EC 3.6.1.31) (PRA-PH)).  
OS HIS2 OR HIS2E OR AQ\_1968.  
GN Aquifex aeolicus.  
OC Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.  
OX NCBI\_TaxID=63363;  
RN [1]

```

RP SEQUENCE FROM N.A.
RC STRAIN=VF5;
RX MEDLINE=98196666; PubMed=9537320;
RA Decker G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA Graham D.E., Overbeek R., Sneed M.A., Keller M., Aujay M., Huber R.,
RA Gelman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex
aerolicus."
RL Nature 392:353-358 (1998).
CC -1- CATALYTIC ACTIVITY: 1-(5-phosphoribosyl)-ATP + H(2)O = 1-(5-
phosphoribosyl)-AMP + diphosphate.
CC -1- CATALYTIC ACTIVITY: 1-(5-phosphoribosyl)-AMP + H(2)O = 1-(5-
phosphoribosyl)-5-[(5-phosphoribosylamino)methylideneamino]imidazole-4-carboxamide.
CC -1- PATHWAY: Histidine biosynthesis; second step.
CC -1- PATHWAY: Histidine biosynthesis; third step.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (by similarity).
CC -1- SIMILARITY: In the N-terminal section; belongs to the PRA-CH
family.
CC -1- SIMILARITY: In the C-terminal section; belongs to the PRA-PH
family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE000765; AAC07730.1; -.
CC PIR; H70468; H70468.
CC DR HAMAP; MF_01019; -.
CC DR InterPro; IPR002496; PRA-CH.
CC DR InterPro; IPR008179; PRA-PH.
CC DR InterPro; IPR008178; PRA-PH/CH.
CC DR Pfam; PF01502; PRA-CH; 1.
CC DR Pfam; PF01503; PRA-PH; 1.
CC DR ProDom; PD002610; PRA-CH; 1.
CC DR ProDom; PD002611; PRA-PH/CH; 1.
CC KM Histidine biosynthesis; Multifunctional enzyme; Hydrolase;
KW Complete proteome.
FT DOMAIN 1 116 PHOSPHORIBOSYL-AMP CYCLOHYDROLASE.
FT DOMAIN 117 205 PHOSPHORIBOSYL-ATP PYROPHOSPHOHYDROLASE.
SQ SEQUENCE 205 AA; 23943 MW; 8DB5DAA3261A2197 CRC64;

Query Match 54.2%; Score 39; DB 1; Length 205;
Best Local Similarity 50.0%; Pred. No. 6.7;
Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 2 IKKRTSGYSHY 13
| : | : | : |
DB 35 IKKTLEGTGYAHY 46

RESULT 6
LYCV_BPPH2 STANDARD; PRT; 258 AA.
ID LYCV_BPPH2
AC P1187;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Lysozyme (EC 3.2.1.17) (Lysis protein) (Muremidase) (Endolysin)
DE (Morphogenesis protein 2).
GN 15.
OS Bacteriophage phi-29.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Podoviridae;
OC phi-29-like viruses.
OX NCBI_TaxID=10756;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87117505; PubMed=3027653;
RA Garvey K.J., Saeedi M.S., Ito J.;

```

```

RT "Nucleotide sequence of Bacillus phage phi 29 genes 14 and 15:
RT homology of gene 15 with other phage lysozymes."
RL Nucleic Acids Res. 14:10001-10008 (1986).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=87106857; PubMed=3803926;
RA Vitek C., Paces V.;
RT "Nucleotide sequence of the late region of Bacillus phage phi 29
RT completes the 19,285-bp sequence of phi 29 genome. Comparison with
RT the homologous sequence of phage PZA."
RL Gene 46:215-225 (1986).
RN [3]
RP SEQUENCE OF 116-258 FROM N.A.
RX MEDLINE=86165873; PubMed=3879485;
RA Garvey K.J., Saeedi M.S., Ito J.;
RT "The complete sequence of Bacillus phage phi 29 gene 16: a protein
RT required for the genome encapsidation reaction."
RL Gene 40:311-316 (1985).
CC -1- FUNCTION: Helps to release the mature phage particles from the
CC cell wall by breaking down the peptidoglycan.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of the 1,4-beta-linkages between N-
CC acetyl-D-glucosamine and N-acetylmuramic acid in peptidoglycan
CC heteropolymers of the prokaryotes cell walls.
CC -1- DOMAIN: LysM repeats are thought to be involved in peptidoglycan
CC binding.
CC -1- SIMILARITY: Belongs to family 24 of glycosyl hydrolases.
CC -1- SIMILARITY: Contains 2 LysM repeats.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X04962; CA28632.1; -.
CC DR EMBL; M14782; AAA32288.1; -.
CC DR EMBL; M14431; AAA88347.1; -.
CC PIR; B24721; MMBPP9.
CC DR HSSP; P00720; 103L.
CC DR InterPro; IPR002186; Glyco_hydro_24.
CC DR InterPro; IPR002482; LysM.
CC DR Pfam; PF01476; LysM; 2.
CC DR Pfam; PF00959; Phage_Lysozyme; 1.
CC DR SMART; SM00257; LysM; 2.
CC KM Late protein; Hydrolase; Glycosidase; Bacteriolytic enzyme; Repeat.
FT ACT SITE 15 15 PROTON DONOR (BY SIMILARITY).
FT REPEAT 163 207 LYSM 1.
FT REPEAT 214 258 LYSM 2.
SQ SEQUENCE 258 AA; 28054 MW; 26AE0C1D927B42D8 CRC64;

Query Match 54.2%; Score 39; DB 1; Length 258;
Best Local Similarity 60.0%; Pred. No. 8.5;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 4 KRTSGYSHY 13
| : | : | : |
DB 29 KHTTIGYGHY 38

RESULT 7
LYCV_BPPZA STANDARD; PRT; 258 AA.
ID LYCV_BPPZA
AC P07540;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-APR-1988 (Rel. 07, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Lysozyme (EC 3.2.1.17) (Lysis protein) (Muremidase) (Endolysin)
DE (late protein gp15).
GN 15.
OS Bacteriophage PZA.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Podoviridae;

```

```

OC phi-29-like viruses.
OX NCBI_TaxID=10757;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87031573; PubMed=1095186;
RA Paces V., Vitek C., Urbanek P.;
RT "Nucleotide sequence of the late region of Bacillus subtilis phage
  P2A, a close relative of phi 29.";
RL Gene 44:107-114(1986).
CC -1- FUNCTION: Helps to release the mature phage particles from the
CC cell wall by breaking down the peptidoglycan.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of the 1,4-beta-linkages between N-
CC acetyl-D-glucosamine and N-acetylmuramic acid in peptidoglycan
CC heteropolymers of the prokaryotic cell walls
CC -1- DOMAIN: LysM repeats are thought to be involved in peptidoglycan
CC binding.
CC -1- SIMILARITY: Belongs to family 24 of glycosyl hydrolases.
CC -1- SIMILARITY: Contains 2 LysM repeats.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M1813; AAA88492.1; -
DR FIK; A26215; WMBP15.
DR HSBP; P00720; 103L.
DR InterPro; IPR002196; Glyco_hydro_24.
DR InterPro; IPR002482; LysM.
DR Pfam; PF01476; LysM; 2.
DR Pfam; PF00959; Phage_lysozyme; 1.
DR SMART; SM00257; LysM; 2.
DR Late protein; Hydrolyase; Glycosylase; Bacteriolytic enzyme; Repeat.
KW ACT_SITE 15 15 PROTON DONOR (BY SIMILARITY).
FT REPEAT 163 207 LysM 1.
FT REPEAT 214 258 LysM 2.
SQ SEQUENCE 258 AA; 28052 MW; 1A57FC534637169 CRC64;

Query Match 54.2%; Score 39; DB 1; Length 258;
Best Local Similarity 60.0%; Pred. No. 8.5;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 4 KRFTSGSHY 13
DB 29 KHTTIGYHY 38

RESULT 8
ARGC STAEF STANDARD; PRT; 341 AA.
ID ARGCS STAEF
AC Q8CP35;
DT 15-MAR-2004 (Rel. 43, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE N-acetyl-gamma-glutamyl-phosphate reductase (NC 1.2.1.38) (AGPR) (N-
DE acetyl-glutamate semialdehyde dehydrogenase)
GN ARGC OR S1212.
OS Staphylococcus epidermidis.
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=1282;
CX [1]
RN SEQUENCE FROM N.A.
RP STRAIN=ATCC 12228;
RX PubMed=12950322;
RA Zhang Y.-Q., Ren S.-X., Li H.-L., Wang Y.-X., Fu G., Yang J.,
RA Qiu Z.-Q., Mao Y.-G., Wang W.-Y., Chen R.-S., Shen Y., Chen Z.,
RA Yuan Z.-H., Zhao G.-P., Qu D., Danchin A., Men Y.-M.;
RT "Genome-based analysis of virulence genes in a non-biofilm-forming
RT Staphylococcus epidermidis strain (ATCC 12228).";
Mol. Microbiol. 49:1577-1593(2003).

```

```

CC -1- CATALYTIC ACTIVITY: N-acetyl-L-glutamate 5-semialdehyde + NADP(+)
CC + phosphate = N-acetyl-5-glutamyl phosphate + NADPH.
CC -1- PATHWAY: Arginine biosynthesis; third step.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -1- SIMILARITY: Belongs to the NagSA dehydrogenase family. Subfamily
CC 1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE016748; AA004811.1; -
DR HAMAP; MF_00150; -; 1.
DR InterPro; IPR000706; AGPR_act_site.
DR InterPro; IPR000534; Semialdh_dh.
DR Pfam; PF01118; Semialdehyde_dh; 1.
DR Pfam; PF02774; Semialdehyde_dhc; 1.
DR ProDom; PD003765; AGPR_act_site; 1.
DR TIGRFAMs; TIGR01850; atgC; 1.
DR PROSITE; PS01224; ARGC; 1.
KW Arginine biosynthesis; Oxidoreductase; NADP; Complete proteome.
FT ACT_SITE 147 147 BY SIMILARITY.
SQ SEQUENCE 341 AA; 38227 MW; CA4B33D3D12DBCBD CRC64;

Query Match 54.2%; Score 39; DB 1; Length 341;
Best Local Similarity 72.7%; Pred. No. 11;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 DIRKFTSGYS 11
DB 120 DIRKFTYGLS 130

RESULT 9
TORZ HAEIN STANDARD; PRT; 825 AA.
ID TORZ HAEIN
AC P44758;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Trimechylamine-N-oxide reductase precursor (EC 1.7.2.3) (TMAO
DE reductase) (Trimechylamine oxidase).
GN TORZ OR BISC OR H10643.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=727;
CX [1]
RN SEQUENCE FROM N.A.
RP STRAIN=RD / KM20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.U., Tomb J.-F., Dougherty B.A., Wetrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uitterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Rhine L.D., Fritchman J.L., Fuhrmann J.L., Geoghegan N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT Rd.";
RL Science 269:496-512(1995).
CC -1- FUNCTION: Reduces trimechylamine-N-oxide (TMAO) into
CC trimechylamine; an anaerobic reaction coupled to energy-yielding
CC reactions (By similarity).
CC -1- CATALYTIC ACTIVITY: Trimechylamine + 2 (ferricytochrome c)-subunit
CC + H(2)O = trimechylamine-N-oxide + 2 (ferrocyclochrome c)-subunit.
CC -1- COFACTOR: Molybdenum (molybdopterin) (By similarity).

```

```

CC -1- SUBCELLULAR LOCATION: Periplasmic (By similarity).
CC -1- SIMILARITY: Belongs to the prokaryotic molybdopterin-containing
CC oxidoreductase family.
CC -1- CAUTION: Was originally (Ref.1) assigned to be a biotin sulfoxide
CC reductase hence the original gene designation of birc.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL: U32747; AAC22303.1; -.
DR PIR: H64083; H64083.
DR HSSP: O57366; 1E01.
DR TIGR: H10643; -.
DR InterPro: IPR009010; Asp decarb_fold.
DR InterPro: IPR006658; B1cG.
DR InterPro: IPR006657; Mol dinuc bind.
DR InterPro: IPR006656; Molybdopterin.
DR InterPro: IPR006655; Prok Mboxed.
DR InterPro: IPR006311; Tat.
DR Pfam: PF00384; molybdopterin; 1.
DR Pfam: PF01568; molybdopterin; 1.
DR TIGR: TIGR01409; Tat signal; seq; 1.
DR TIGR: TIGR01409; Tat signal; seq; 1.
DR PROSITE: PS00551; MOLYBDOPTERIN_PROK_1; FALSE_NEG.
DR PROSITE: PS00490; MOLYBDOPTERIN_PROK_2; 1.
DR PROSITE: PS00932; MOLYBDOPTERIN_PROK_3; 1.
DR Oxidoreductase; Molybdenum; Periplasmic; Signal; Complete proteome.
KW SIGNAL
FT CHAIN 1 825 TRIMETHYLAMINE-N-OXIDE REDUCTASE.
FT SEQUENCE 825 AA; 91053 MW; 4254287D42B45E55 CRC64;
SQ
Query Match 54.2%; Score 39; DB 1; Length 825;
Best Local Similarity 46.2%; Pred. No. 29;
Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;
QY 1 DIRKRTSGSYSHY 13
DB 303 DFLKRTSGYAKF 315

```

```

CC -1- SIMILARITY: Contains 1 HMG box domain.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL: X98884; CAA67387.1; -.
DR PIR: T30334; T30334.
DR HSSP: O05783; 1HMA.
DR InterPro: IPR000910; HMG_12_box.
DR InterPro: IPR001680; WD40.
DR Pfam: PF00505; HMG_box; 1.
DR Pfam: PF00400; WD40; 4.
DR ProDom: PD000018; WD40; 1.
DR SMART: SM00398; HMG; 1.
DR SMART: SM00320; WD40; 4.
DR PROSITE: PS50118; HMG_BOX_2; 1.
DR PROSITE: PS00678; WD_REPEATS_1; 2.
DR PROSITE: PS50082; WD_REPEATS_2; 3.
DR PROSITE: PS50294; WD_REPEATS_REGION; 1.
DR Nucleic protein; DNA-binding; Repeat; WD repeat.
KW Repeat
FT REPEAT 11 50 WD 1.
FT REPEAT 52 91 WD 2.
FT REPEAT 92 131 WD 3.
FT REPEAT 134 173 WD 4.
FT REPEAT 184 223 WD 5.
FT REPEAT 228 267 WD 6.
FT REPEAT 271 310 WD 7.
FT DOMAIN 821 859 ASP/GDU-RICH (ACIDIC).
FT DOMAIN 859 878 GLU-RICH (ACIDIC).
FT DNA BIND 1013 1076 HMG-BOX.
FT SEQUENCE 1127 AA; 124649 MW; 4F4CA33C7422082 CRC64;
SQ
Query Match 54.2%; Score 39; DB 1; Length 1127;
Best Local Similarity 63.6%; Pred. No. 4;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
QY 1 DIRKRTSGSYSHY 11
DB 829 DFRSLTLAGYS 839

```

```

RESULT 10
AND1_XENLA
ID AND1_XENLA STANDARD; PRT: 1127 AA.
AC 013046;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Acidic nucleoplasmic DNA-binding protein 1 (And-1).
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A., SUBCELLULAR LOCATION, SUBUNIT, AND DNA BINDING.
RC TISSUE=Oocyte;
RX MEDLINE=97318764; PubMed=9175701;
RA Koehler A., Schmidt-Zachmann M.S., Franke W.W.;
RT "AND-1, a natural chimeric DNA-binding protein, combines an HMG-box
RT with regulatory WD-repeats."
CC -1- FUNCTION: Binds DNA with high affinity. May also be involved in
CC protein-protein interactions.
CC -1- SUBUNIT: Homodimer (Probable).
CC -1- SUBCELLULAR LOCATION: Nucleolus; nucleoplasm. Transiently
CC cytoplasmic during mitosis.
CC -1- TISSUE SPECIFICITY: Found in oocytes and in various other cells.
CC -1- SIMILARITY: Contains 7 WD repeats.

```

```

RESULT 11
HIS3_METKA
ID HIS3_METKA STANDARD; PRT: 141 AA.
AC P58836;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Phosphoribosyl-AMP cyclohydrolase (EC 3.5.4.19) (PRA-CH).
GN HIS1 OR HIS1.2 OR MK1537.
OS Archaea; Euryarchaeota; Methanopyrti; Methanopyrales; Methanopyraceae;
OC Archaea; Euryarchaeota; Methanopyrti; Methanopyrales; Methanopyraceae;
OX NCBI_TaxID=2320;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AV19 / DSM 6324 / JCM 9639;
RX MEDLINE=21927647; PubMed=11930014;
RA Slesarev A.I., Mezheva K.V., Makarova K.S., Polushin N.N.,
RA Shcherbina O.V., Shakhova V.V., Belova G.I., Aravind L.,
RA Natale D.A., Rogozin I.B., Tatusov R.L., Wolf Y.I., Stetter K.O.,
RA Malyshev A.G., Koonin E.V., Kozlovskiy S.A.;
RT "The complete genome of hyperthermophile Methanopyrus kandleri AV19
RT and monophyly of archaeal methanogens."
RL Proc. Natl. Acad. Sci. U.S.A. 99:4644-4649(2002).
CC -1- CATALYTIC ACTIVITY: 1-(5-phosphoribosyl)-AMP + H(2)O = 1-(5-
CC phosphoribosyl)-5-[(5-
CC phosphoribosylamino)methylideneamino]imidazole-4-carboxamide.

```

```

CC -1- PATHWAY: Histidine biosynthesis; third step.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: Belongs to the PRA-CH family.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; AB010444; AAM02750.1; -.
CC HAMAP; MF_01021; -.
CC InterPro; IPR002496; PRA-CH.
CC Pfam; PF01502; PRA-CH; 1.
CC ProDom; PD002610; PRA-CH; 1.
CC Histidine biosynthesis; Hydrolyase; Complete proteome.
CC SEQUENCE 141 AA; 16089 MM; 14E92D9A4DPAF87 CRC64;
-----
QY Query Match 51.4%; Score 37; DB 1; Length 141;
Best Local Similarity 41.7%; Pred. No. 10;
Matches 5; Conservative 4; Mismatches 3; Indels 0; Gaps 0;
-----
Db 2 IRKFTSGYSHY 13
53 YRRLTIGYAHY 64
-----
RESULT 12
LYCV_BPB03 STANDARD; PRT; 263 AA.
AC Q37896; 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Lysozyme (EC 3.2.1.17) (Lysis protein) (Muramidase) (Endolysin)
DE (Morphogenesis protein 2).
GN 15.
OS Bacteriophage B103.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Podoviridae;
OC phi-29-like viruses.
OC NCBI_Taxid=10778;
OX NCB1_Taxid=10778;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=98019084; PubMed=9358052;
RA Pecenkova T., Benes V., Paces J., Vicek C., Paces V.;
RA "Bacteriophage B103: complete DNA sequence of its genome and
RA relationship to other Bacillus phages.";
RL Gene 199:157-163(1997).
CC -1- FUNCTION: Helps to release the mature phage particles from the
CC cell wall by breaking down the peptidoglycan (By similarity).
CC -1- CATALYTIC ACTIVITY: Hydrolysis of the 1,4-beta-linkages between N-
CC acetyl-D-glucosamine and N-acetylmuramic acid in peptidoglycan
CC heteropolymers of the prokaryotes cell walls.
CC -1- DOMAIN: Lyso repeats are thought to be involved in peptidoglycan
CC binding.
CC -1- SIMILARITY: Belongs to family 24 of glycosyl hydrolases.
CC -1- SIMILARITY: Contains 2 Lyso repeats.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; X99260; CA67646.1; -.
CC HSSP; P00720; 103L.
CC InterPro; IPR002196; Glyco_hydro_24.
CC InterPro; IPR002482; LysoM.
CC Pfam; PF01476; LysoM; 2.

```

```

DR Pfam; PF00959; Phage_lysozyme; 1.
DR SMART; SM00257; LysoM; 2.
KW Late protein; Hydrolyase; Glycosidase; Bacteriolytic enzyme; Repeat.
FT ACT SITE 15 15 PROTON DONOR (BY SIMILARITY).
FT REPEAT 165 209 LYSO 1.
FT REPEAT 219 263 LYSO 2.
SQ SEQUENCE 263 AA; 29264 MM; 5C1A7C8282ADBCE CRC64;
-----
QY Query Match 51.4%; Score 37; DB 1; Length 263;
Best Local Similarity 60.0%; Pred. No. 20;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
-----
Db 4 KRFTSGYSHY 13
29 KYRTIGYGHY 38
-----
RESULT 13
SER1_DROME STANDARD; PRT; 265 AA.
ID SER1_DROME Q9VAD9; Q9VAD9;
AC P17205; Q9VAD9; 01-AUG-1990 (Rel. 15, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Serine proteases 1/2 precursor (EC 3.4.21.-).
GN (SER99DA OR SER1 OR CG7877) AND (SER99DB OR SER2 OR CG15519).
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OC NCBI_Taxid=7227;
OX NCB1_Taxid=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=89219063; PubMed=2469005;
RA Yun Y., Davis R.L.;
RA "Levels of RNA from a family of putative serine protease genes are
RA reduced in Drosophila melanogaster dunce mutants and are regulated by
RA cyclic AMP.";
RL Mol. Cell. Biol. 9:692-700(1989).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkely;
RC MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner S.N.,
RA Sutton G.G., Wortman T.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brannon R.C., Rogers Y.H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., McKlos G.L.G.,
RA Abrial J.F., Agapayni A., An H.-J., Andrews-Pfankoch C., Baldwin D.,
RA Ballaw R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Bortchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Danke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dierz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fessler C., Gabrielian A.B., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegam C.,
RA Jalali M., Kalish F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kammel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Morklov G., Mlshina N.V., Moberly C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo W., Pittman G.S., Pan S., Pollard Y., Puti V., Reese M.G.,
RA Rainert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,

```

```

RA Svitak R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Massaman D.A., Weinstock G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
CC -1- FUNCTION: Major function may be to aid in digestion.
CC -1- TISSUE SPECIFICITY: Abundantly expressed in the larval gut.
CC -1- DEVELOPMENTAL STAGE: Began to appear at late embryo stage and
CC continued to increase in abundance throughout the larval stage.
CC They are not present in pupae but reappeared in the adult.
CC -1- MISCELLANEOUS: The sequence shown is that of SER1.
CC -1- SIMILARITY: Belongs to peptidase family S1.
CC -1- CAUTION: Ref.2 sequence differs from that shown due to erroneous
CC gene model prediction.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; M24379; AAB02552.1; -
DR EMBL; M24379; AAB02553.1; -
DR EMBL; AE003771; AAN14201.1; -
DR EMBL; AE003771; AAF56972.1; -
DR PIR; A38894; A38894.
DR PIR; J50260; J50260.
DR HSSP; P00761; 1EPT.
DR MEROPS; S01.00A; -.
DR FLYBASE; FBgn0003357; Ser99da.
DR FLYBASE; FBgn0003357; Ser99db.
DR InterPro; IPR009003; Cys_ser_trypsin.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001314; Peptidase_S1A.
DR InterPro; IPR008256; Peptidase_S1B_V8.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PRO0722; CHWOTRPSIN.
DR PRINTS; PRO0835; V8PROTASE.
DR SMART; SM00020; TRYP_SPE; 1.
DR PROSITE; PS00240; TRYPsin_DOM; 1.
DR PROSITE; PS00134; TRYPsin_HIS; 1.
DR PROSITE; PS00135; TRYPsin_SER; 1.
KW Hydrolase; Serine protease; Signal; Zymogen; Multigene family.
FT SIGNAL 1 21 PROBABLE.
FT PROPEP 22 35 BY SIMILARITY.
FT CHAIN 36 265 SERINE PROTEASES 1/2.
FT ACT_SITE 78 78 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 123 123 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 215 215 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 63 79 BY SIMILARITY.
FT DISULFID 189 201 BY SIMILARITY.
FT DISULFID 211 239 BY SIMILARITY.
FT CONFLICT 14 14 A -> T (IN REF. 1; AAB02553).
SQ SEQUENCE 265 AA; 28469 MW; 44848C523F03384B CRC64;

Query Match 51.4%; Score 37; DB 1; Length 265;
Best Local Similarity 53.8%; Pred. No. 20;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

```

```

DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Possible protein kinase B12 (EC 2.7.1.-).
GN B12R.
OS Vaccinia virus (strain Copenhagen).
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC Orthopoxvirus.
OC NCBI_TaxID=10249;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91021027; PubMed=2219722;
RA Geibel S.J., Johnson G.P., Perkus M.E., Davis S.W., Winslow J.P.,
RA Paolletti E.;
RT "The complete DNA sequence of vaccinia virus.";
RL Virology 179:247-266(1990).
RN [2]
RP COMPLETE GENOME.
RA Geibel S.J., Johnson G.P., Perkus M.E., Davis S.W., Winslow J.P.,
RA Paolletti E.;
RT "Appendix to 'The complete DNA sequence of vaccinia virus.'";
RL Virology 179:517-563(1990).
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC POXVIRUSES SUBFAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; M35027; AAA48209.1; -
DR PIR; B42527; TVVZB2.
DR InterPro; IPR000719; Prot_kinase.
DR InterPro; IPR008271; Ser_thr_pkin_AS.
DR ProDom; PD000001; Prot_Kinase; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; FALSE_NEG.
DR PROSITE; PS00108; PROTEIN KINASE ST; FALSE_NEG.
DR PROSITE; PS00111; PROTEIN KINASE_DOM; 1.
KW Transferase; Serine/threonine-protein kinase; ATP-binding.
FT DOMAIN 1 283
FT SEQUENCE 283 AA; 33366 MW; 3D61A894F4189E2A CRC64;

Query Match 51.4%; Score 37; DB 1; Length 283;
Best Local Similarity 50.0%; Pred. No. 22;
Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

```

```

QY 1 DIRKFTSGYSHY 13
DB 31 DIQGRITNGYPAY 43

RESULT 14
KRB2_VACCC STANDARD; PRT; 283 AA.
ID KRB2_VACCC
AC P21098;

```

```

QY 2 IRKFTSGYSHY 13
DB 31 VRKFTSGYSHY 42

RESULT 15
KRB2_VACCV STANDARD; PRT; 283 AA.
AC P24352;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Possible protein kinase B12 (EC 2.7.1.-).
GN B12R.
OS Vaccinia virus (strain WR).
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC Orthopoxvirus.
OC NCBI_TaxID=10254;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91259063; PubMed=2045793;
RA Smith G.L., Chan Y.S., Howard S.T.;
RT "Nucleotide sequence of 42 kbp of vaccinia virus strain WR from near
the right inverted terminal repeat.";

```



```

RL U. Gen. Virol. 72:1349-1376(1991).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=90111697; PubMed=2607336;
RA Howard S.T., Smith G.L.;
RT "Two early vaccinia virus genes encode polypeptides related to
RT protein kinases";
RL U. Gen. Virol. 70:3187-3201(1989).
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC POXVIRUS SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D11079; BAA01842.1; -
DR EMBL; D00629; BAA00520.1; -
DR PIR; A33610; TVVZBK.
DR InterPro; IPR000719; Prot_kinase.
DR InterPro; IPR008271; Ser_thr_pkin_AS.
DR ProDom; PD000001; Prot_kinase; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; FALSE_NEG.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; FALSE_NEG.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1
DR Transferase; Serine/threonine-protein kinase; ATP-binding.
KM DOMAIN 1 283
FT SEQUENCE 283 AA; 33310 MW; D2819B94F73F9C77 CRC64;
SQ

```

```

Query Match 51.4%; Score 37; DB 1; Length 283;
Best Local Similarity 50.0%; Pred. No. 22;
Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

```

```

QY 2 IRKRFTSGVSHY 13
   :|:|:|:|:|:|
Db 31 VRKNFTSFFVNY 42

```

Search completed: October 5, 2004, 08:01:53  
 Job time : 3.16358 secs

This Page Blank (uspio)

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2004, 07:34:11 : Search time 6.8817 seconds  
(without alignments)  
596.081 Million cell updates/sec

Title: US-09-805-290A-17  
Perfect score: 72  
Sequence: 1 DIRKFRSGSHY 13  
Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL\_25:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mmc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriophage:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	43	59.7	Q86DC2	Q86DC2 caenorhabdi
2	43	59.7	Q86DC2	Q86DC2 caenorhabdi
3	43	59.7	Q86DC2	Q86DC2 caenorhabdi
4	42	58.3	Q86DC2	Q86DC2 caenorhabdi
5	42	58.3	Q86DC2	Q86DC2 caenorhabdi
6	41	56.9	Q86DC2	Q86DC2 caenorhabdi
7	40	55.6	Q86DC2	Q86DC2 caenorhabdi
8	40	55.6	Q86DC2	Q86DC2 caenorhabdi
9	40	55.6	Q86DC2	Q86DC2 caenorhabdi
10	40	55.6	Q86DC2	Q86DC2 caenorhabdi
11	40	55.6	Q86DC2	Q86DC2 caenorhabdi
12	40	55.6	Q86DC2	Q86DC2 caenorhabdi
13	40	55.6	Q86DC2	Q86DC2 caenorhabdi
14	40	55.6	Q86DC2	Q86DC2 caenorhabdi
15	40	55.6	Q86DC2	Q86DC2 caenorhabdi
16	40	55.6	Q86DC2	Q86DC2 caenorhabdi

17	39.5	54.9	429	5	Q86P4	Q86P4 caenorhabdi
18	39	54.2	195	5	Q21813	Q21813 caenorhabdi
19	39	54.2	249	2	Q83VN3	Q83VN3 streptococc
20	39	54.2	249	2	Q83VN0	Q83VN0 streptococc
21	39	54.2	249	2	Q83VN7	Q83VN7 streptococc
22	39	54.2	258	9	Q84ML9	Q84ML9 bacillus am
23	39	54.2	337	16	Q92XG1	Q92XG1 initobium m
24	39	54.2	341	16	Q8CP35	Q8CP35 staphylococ
25	39	54.2	396	13	Q7ZT10	Q7ZT10 xenopus lae
26	39	54.2	418	16	Q8NXX1	Q8NXX1 staphylococ
27	39	54.2	418	16	Q8NXX2	Q8NXX2 staphylococ
28	39	54.2	419	2	Q8KUC5	Q8KUC5 staphylococ
29	39	54.2	431	16	Q8XB68	Q8XB68 staphylococ
30	39	54.2	530	16	Q83BQ4	Q83BQ4 enterococcu
31	39	54.2	758	17	Q9UY62	Q9UY62 pyrococcus
32	39	54.2	945	5	Q814Y8	Q814Y8 plasmodium
33	39	54.2	1132	10	Q8SOP3	Q8SOP3 oryza sativ
34	39	54.2	1253	10	Q8SAQ4	Q8SAQ4 arabidopsis
35	39	54.2	1786	13	Q7ZT14	Q7ZT14 poephilla gu
36	39	54.2	3763	5	Q812A1	Q812A1 dictyosteli
37	38.5	53.5	284	16	Q97GQ2	Q97GQ2 clostridium
38	38	52.8	250	16	Q881H7	Q881H7 pseudomonas
39	38	52.8	300	16	Q9PR26	Q9PR26 ureaplasma
40	38	52.8	316	11	Q922H9	Q922H9 mus musculu
41	38	52.8	317	11	Q70524	Q70524 cricetus
42	38	52.8	320	4	Q91352	Q91352 homo sapien
43	38	52.8	320	4	Q8KSR7	Q8KSR7 salmonella
44	38	52.8	349	2	Q9X388	Q9X388 bacillus an
45	38	52.8	354	2	Q848A2	Q848A2 clostridium

## ALIGNMENTS

RESULT 1  
Q86DC2 ID Q86DC2 PRELIMINARY; PRT; 449 AA.  
AC Q86DC2; 01-JUN-2003 (TRENBLREI.. 24, Created)  
DT 01-JUN-2003 (TRENBLREI.. 24, Last sequence update)  
DT 01-OCT-2003 (TRENBLREI.. 25, Last annotation update)  
DE Hypothetical protein R04A9.5.  
GN R04A9.5.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Pelodermata; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RX MEDLINE=99069613; PubMed=9851916;  
RA Waterston R.;  
RT "Genome sequence of the nematode C. elegans: a platform for  
investigating biology. The C. elegans Sequencing Consortium.";  
RL Science 282:2012-2018(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RX Gelsel C.;  
RT "The sequence of C. elegans coemid R04A9.";  
RL Submitted (DEC-1995) to the EMBL/Genbank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RX Submitted (May-2003) to the EMBL/Genbank/DBJ databases.  
RL EMBL; U41550; AAP31434.1;  
DR WormPep; R04A9.5b; CE33932.  
DR GO; GO:0005634; C:nucleus; IEA.  
DR GO; GO:0003700; F:transcription factor activity; IEA.  
DR GO; GO:0006355; P:regulation of transcription; DNA-dependent; IEA.  
DR InterPro; IPR001356; Homeobox.  
DR Pfam; PF00046; homeobox; 1.

DR ProDom; PD000010; Homeobox; 1.  
 DR SMART; SM00389; HOX; 2.  
 DR PROSITE; PS50071; HOMEBOX\_2; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 449 AA; 52190 MW; ED894CCE1961F614 CRC64;

Query Match 59.7%; Score 43; DB 5; Length 449;  
 Best Local Similarity 80.0%; Pred. No. 29;  
 Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 KRFTSGYSH 12  
 Db 291 KRFTLGIVH 300

## RESULT 2

ID 067406 PRELIMINARY; PRT; 759 AA.  
 AC 067406;  
 DT 01-AUG-1998 (TREMBLrel. 07, Created)  
 DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)  
 DE 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
 DE Cellulose synthase catalytic subunit.  
 GN BCSA OR AQ.1407.  
 OS Aquifex aeolicus.  
 OC Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.  
 CX NCBI\_TaxId=63363;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=VF5;  
 RA MEDLINE=98196666; PubMed=9537320;  
 RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,  
 RA Graham D.E., Overbeek R., Shead M.A., Keller M., Anjay M., Huber R.,  
 RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;  
 RT "The complete genome of the hyperthermophilic bacterium Aquifex  
 aeolicus.";  
 RT Nature 392:353-358 (1998).  
 DR EMBL; AE000738; AAC07360.1; -.  
 DR PIR; D70422; D70422.  
 DR GO; GO:0016020; Cmembrane; IEA.  
 DR GO; GO:0016760; P:cellulose synthase (UDP-forming) activity; IEA.  
 DR GO; GO:0030244; P:cellulose biosynthesis; IEA.  
 DR GO; GO:0006011; P:UDP-glucose metabolism; IEA.  
 DR InterPro; IPR005150; Cellulose synth.  
 DR InterPro; IPR003919; Cell synth A.  
 DR InterPro; IPR001173; Glyco trans.2.  
 DR Pfam; PF03552; Cellulose\_synth; 1.  
 DR Pfam; PF00535; Glycos\_transf\_2; 1.  
 DR PRINTS; PR01439; CELLSTHASEA.  
 KW COMPLETE proteome.  
 SQ SEQUENCE 759 AA; 89561 MW; A71C7333F015082C CRC64;

Query Match 59.7%; Score 43; DB 16; Length 759;  
 Best Local Similarity 70.0%; Pred. No. 51;  
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 4 KRFTSGYSH 13  
 Db 65 KRFTGTYHY 74

## RESULT 3

ID 021694 PRELIMINARY; PRT; 879 AA.  
 AC 021694;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-OCT-2001 (TREMBLrel. 18, Last sequence update)  
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
 DE Hypothetical protein.  
 GN R0949.5.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.

OX NCBI\_TaxId=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RX MEDLINE=99069613; PubMed=9851916;  
 RA None;  
 RT "Genome sequence of the nematode C. elegans: a platform for  
 RT investigating biology. The C. elegans Sequencing Consortium.";  
 RT Science 282:2012-2018 (1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RA Gelsel C.;  
 RT "The sequence of C. elegans cosmid R0949.";  
 RT Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.

QY 3 KRFTSGYSH 12  
 Db 721 KRFTLGIVH 730

## RESULT 4

ID 09PNZ2 PRELIMINARY; PRT; 401 AA.  
 AC 09PNZ2;  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)  
 DE Putative integral membrane protein.  
 GN C10941C.  
 OS Campylobacter jejuni.  
 OX NCBI\_TaxId=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RX MEDLINE=99069613; PubMed=9851916;  
 RA None;  
 RT "Genome sequence of the nematode C. elegans: a platform for  
 RT investigating biology. The C. elegans Sequencing Consortium.";  
 RT Science 282:2012-2018 (1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RA Gelsel C.;  
 RT "The sequence of C. elegans cosmid R0949.";  
 RT Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.

Query Match 59.7%; Score 43; DB 5; Length 879;  
 Best Local Similarity 80.0%; Pred. No. 59;  
 Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 KRFTSGYSH 12  
 Db 721 KRFTLGIVH 730

```

OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC Campylobacteraceae; Campylobacter.
OX NCBI_TaxID=197;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCCTC 11168;
RX MEDLINE=20150912; PubMed=10688204;
RA Parkhill J., Wren B.W., Mungall K., Kettle J.M., Churcher C.,
  Basham D., Chillingworth T., Davies R.M., Feltham T., Holtroyd S.,
  Jagsell K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,
  Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,
  Whitehead S., Barrall B.G.;
RT "The genome sequence of the food-borne pathogen Campylobacter jejuni
  reveals hypervariable sequences."
RL Nature 403:665-668(2000).
DR EMBL; AL139076; CAB73198.1; -.
DR PIR; D81368; D81368.
DR GO; GO:0016020; C:membrane; IEA.
DR InterPro; IPR003838; DUF214.
DR Pfam; PF00687; Ptsx; 1.
KM Complete proteome.
SQ SEQUENCE 401 AA; 45030 MW; 23C509B1F8E583EA CRC64;

Query Match 58.3%; Score 42; DB 16; Length 401;
Best Local Similarity 53.8%; Pred. No. 39;
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 DIRKRTSGSYSHY 13
DB 183 DVKARFTSGLAFF 195

RESULT 5
Q86KC9 PRELIMINARY; PRT; 453 AA.
ID Q86KC9
AC Q86KC9;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
OX NCBI_TaxID=44689;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=AX4;
RX MEDLINE=22092622; PubMed=12097910;
RA Glockner G., Eichinger L., Szafarski K., Pacheco J., Dear P.,
  Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,
  Tunggal B., Cox B., Quail M.A., Platzer M., Rosenthal A., Neugeb A.A.;
RT "Sequence and analysis of chromosome 2 of Dictyostelium discoideum."
RL Nature 418:79-85(2002).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=AX4;
RA Baumgart C.;
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
KW Hypothetical protein.
SQ SEQUENCE 453 AA; 51421 MW; 5C79A3D45C13E1CD CRC64;

Query Match 58.3%; Score 42; DB 5; Length 453;
Best Local Similarity 72.7%; Pred. No. 44;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 TRKRTSGSYSHY 12
DB 213 IIVKFGSYSHY 223

RESULT 6
Q88AW5 PRELIMINARY; PRT; 877 AA.
ID Q88AW5

```

```

AC Q88AW5;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE TonB system transport protein, putative.
GN PSPT0268.
OS Pseudomonas syringae (pv. tomato).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=323;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=DC3000;
RA Buell R., Joardar V., Khouri H., Fedorova N., Tran B., Russell D.,
  Berry K., Utecherback T., Van Aken S., Feldblyum T., Gwin M.,
  Dodson R., Deboy R., Durkin A., Kolonay J., Madupu R., Daugherty S.,
  Brihac L., Beanan M., Haft D., Selengut J., Nelson W., Davidson T.,
  White O., Fraser C., Collier A.;
RT "Complete sequence of Pseudomonas syringae."
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE016856; AA053814.1; -.
DR TIGR; PSPT0268; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0004872; F:receptor activity; IEA.
DR GO; GO:0005215; F:transporter activity; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF00593; TonB_dep_Rec; 1.
KM Complete proteome.
SQ SEQUENCE 877 AA; 93959 MW; 78697DD5FE94835 CRC64;

Query Match 56.9%; Score 41; DB 16; Length 877;
Best Local Similarity 58.3%; Pred. No. 14+02;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 TRKRTSGSYSHY 13
DB 793 LTRKRTSGYNDY 804

RESULT 7
Q8D4X1 PRELIMINARY; PRT; 168 AA.
ID Q8D4X1
AC Q8D4X1;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Chemotaxis signal transduction protein.
GN VW21167.
OS Vibrio vulnificus.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=672;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=CMCP6;
RA Rhee J.H., Kim S.Y., Chung S.S., Kim J.J., Moon Y.H., Jeong H.,
  Choy H.E.;
RT "Complete genome sequence of Vibrio vulnificus CMCP6."
RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE016812; AA008065.1; -.
DR GO; GO:0005622; C:intracellular; IEA.
DR GO; GO:0004871; F:signal transducer activity; IEA.
DR GO; GO:0006935; P:chemotaxis; IEA.
DR GO; GO:0007165; P:signal transduction; IEA.
DR InterPro; IPR002545; Chem.
DR Pfam; PF01584; Chem; 1.
DR SMART; SM00260; Chem; 1.
DR PROSITE; PS50851; CHEW; 1.
KW Complete proteome.
SQ SEQUENCE 168 AA; 18776 MW; 9AA813E8891B7B61 CRC64;

Query Match 55.6%; Score 40; DB 16; Length 168;

```

Best Local Similarity 53.8%; Pred. No. 36;  
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;  
QY 1 DIRKRTSGYSHY 13  
Db 76 DLRKRCVGECHY 88

RESULT 8  
ID 08UCCK3 PRELIMINARY; PRT; 250 AA.  
AC 08UCCK3;  
DT 01-JUN-2002 (TREMBlrel. 21, Created)  
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)  
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
DE Lipase esterase.  
GN ATU2481 OR AGR C.4503.  
OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
OC Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.  
OX NCBI\_TaxID=176299;  
RN (1)  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21608550; PubMed=11743193;  
RA Wood D.W., Seubald J.C., Kaul R., Monks D.E., Kitajima J.P.,  
RA Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F., Jr., Moo L.,  
RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Bovee D. St.,  
RA Chapman P., Clendinning J., Deatherage G., Gillet M., Grant C.,  
RA Kutyavin T., Levy R., Li M.-J., McClelland E., Palmeri A.,  
RA Raymond C., Rouse G., Saenphitachak C., Wu Z., Romero P., Gordon D.,  
RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,  
RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,  
RA Chunley F., Tingey S.V., Tomb J.F., Gordon M.P., Olson M.V.,  
RA Nester E.W.,  
RT "The genome of the natural genetic engineer Agrobacterium tumefaciens  
C58.";  
RT Science 294:2317-2323(2001).  
RL [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21608551; PubMed=11743194;  
RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,  
RA Quicler B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,  
RA Houtell K., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu F.,  
RA Wollam C., Allinger M., Doughy D., Scott C., Lappas C., Markelz B.,  
RA Flanagan C., Crowell C., Garson J., Lomo C., Sear C., Strub G.,  
RA Cielo C., Slater S.;  
RT "Genome sequence of the plant pathogen and biotechnology agent  
Agrobacterium tumefaciens C58.";  
RT Science 294:2323-2328(2001).  
RL EMBL; AE009195; AAL43468.1;  
DR EMBL; AE008161; AAK88215.1;  
DR PIR; AF2881; AF2881.  
DR PIR; F97657; F97657.  
DR GO; GO:0003824; P: catalytic activity; IEA.  
DR InterPro; IPR000379; Ser\_ester.  
KW Complete proteome.  
SQ SEQUENCE 250 AA; 27086 MW; 1DB50F727BA1FED4 CRC64;  
Query Match 55.6%; Score 40; DB 16; Length 250;  
Best Local Similarity 50.0%; Pred. No. 54;  
Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;  
QY 2 TRKRTSGYSHY 13  
Db 146 MEORWADGMSHY 157

DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)  
DE BL3487 protein.  
GN BL3487  
OS Bradyrhizobium japonicum.  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
OC Bradyrhizobiaceae; Bradyrhizobium.  
OX NCBI\_TaxID=375;  
RN (1)  
RP SEQUENCE FROM N.A.  
RX STRAIN=USDA 110;  
RA Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiumi T.,  
RA Sasamoto S., Watanabe A., Idesawa K., Iriyuchi M., Kawashima K.,  
RA Kohara M., Matsumoto M., Shimo S., Tsuruoka H., Wada T., Yamada M.,  
RA Tabata S.;  
RT "Complete genomic sequence of nitrogen-fixing symbiotic bacterium  
Bradyrhizobium japonicum USDA110.";  
RT DNA Res. 9:189-197(2002).  
DR EMBL; AP005947; BAC48752.1;  
KW Complete proteome.  
SQ SEQUENCE 305 AA; 32572 MW; 30350FDA74E1D7C1 CRC64;  
Query Match 55.6%; Score 40; DB 16; Length 305;  
Best Local Similarity 80.0%; Pred. No. 67;  
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
QY 1 DIRKRTSGY 10  
Db 74 DIRKRTSGY 83

RESULT 10  
ID 08MS21 PRELIMINARY; PRT; 331 AA.  
AC 08MS21;  
DT 01-OCT-2002 (TREMBlrel. 22, Created)  
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)  
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
DE GM14972P (EC 2.1.1.73) (Modification methylase) (Cytosine-specific  
methyltransferase).  
GN MT2 OR CG10692.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachyera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=1227;  
RN (1)  
RP SEQUENCE FROM N.A.  
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,  
RA Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise B.,  
RA George R., Gonzalez M., Guarini H., Krommiller B., Li P., Liao G.,  
RA Miranda A., Mungall C.J., Nunoo U., Pacled J., Paragas V., Park S.,  
RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,  
RA Cealiker S.;  
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.  
CC -1- CATALYTIC ACTIVITY: S-ADENOSYL-L-METHIONINE + DNA CYTOSINE = S-  
ADENOSYL-L-HOMOCYSTEINE + DNA 5-METHYLCYTOSINE.  
CC -1- SIMILARITY: BELONGS TO THE C5-METHYLTRANSFERASE FAMILY.  
DR EMBL; AY118475; AAM49844.1;  
DR FlyBase; FBgn028707; Mt2.  
DR GO; GO:0003677; F:DNA binding; IEA.  
DR GO; GO:0008168; F:methyltransferase activity; IEA.  
DR GO; GO:0008326; F:site-specific DNA-methyltransferase (cytosine . . .) IEA.  
DR GO; GO:0016740; F:transferase activity; IEA.  
DR GO; GO:0006306; P:DNA methylation; IEA.  
DR InterPro; IPR001525; C5\_DNA\_meth.  
DR Pfam; PF00145; DNA\_methylase; 1.  
DR PROSITE; PS00095; C5\_MTASE\_2; 1.  
KW Methyltransferase; Transferase.  
SQ SEQUENCE 331 AA; 37900 MW; 67519BB1514FC64A CRC64;  
Query Match 55.6%; Score 40; DB 5; Length 331;  
Best Local Similarity 75.0%; Pred. No. 73;

Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 6 FTGYSHY 13  
 |||||  
 220 FTGYSHY 227

Db

RESULT 11

Q81P96 PRELIMINARY; PRT; 331 AA.

AC Q81P96; 01-MAR-2003 (TrEMBLrel. 23, Created)

DT 01-MAR-2003 (TrEMBLrel. 25, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE CG10692-P5.

GN M22 OR CG10692.

OS Drosophila melanogaster (fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI\_Taxid=7227;

[1]

RP SEQUENCE FROM N.A.

RX MEDLINE=20196006; PubMed=10731132;

RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,

RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,

RA Brandon R.C., Rogers Y.H., Blazet R.G., Champe M., Pfeiffer B.D.,

RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabot G.L.,

RA Abail J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,

RA Ballew R.M., Baer A., Bayendale J., Bayraktaroglu L., Beasley E.M.,

RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,

RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,

RA Buttle K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,

RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,

RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,

RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,

RA Dublin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,

RA Fodor C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,

RA Glöckler A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,

RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,

RA Hostin D., Houston K.A., Howland T.J., Hernandez J.R., Houck J.,

RA Jatali M., Kalush F., Karpen G.H., Ke Z., Kenison J.A., Ketchum K.A.,

RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

RA Lasako P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,

RA Liu X., Mattei B., McIntosh T.C., McLeod K.P., McPherson D.,

RA Merkulov G., Milhina N.V., Mobarry C., Morris J., Moshrefi A.,

RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,

RA Nelson D.R., Nelson K.A., Nixon K., Nusken D.R., Pacle J.M.,

RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,

RA Reinert K., Remington K., Sanders R.D., Scheeler F., Shen H.,

RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,

RA Spier E., Spreading A.C., Stapleton M., Strong R., Sun E.,

RA Svitek R., Tecor C., Turner R., Venter E., Wang A.H., Wang X.,

RA Wang Z.Y., Wasserman D.A., Weinstein G.M., Weisenbach J.,

RA Williams S.M., Woodgerg, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,

RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,

RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,

RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;

RT "The genome sequence of Drosophila melanogaster.";

RL Science 287:2185-2195 (2000).

[2]

RP SEQUENCE FROM N.A.

RA Celniker S.E., Adams M.D., Krommiller B., Wan K.H., Holt R.A.,

RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,

RA Banon J., An H., Baldwin D., Banon J., Beeson K.Y., Busam D.A.,

RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,

RA Carlson K., Dossert V., Doup L.E., Doyle C., Dresnek D., Farfan D.,

RA Fierle S., Frise E., Galle R.F., Garg N.S., George R.A.,

RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,

RA Ibegam C., Jatali M., Kruse D., Li P., Mattei B., Moshrefi A.,

RA McInosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunco J.,

RA Pacle J., Paragas V., Park S., Patel S., Pfeiffer B., Scheeler F.,

RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scherer F.,

RA Stapleton M., Strong R., Svitek R., Tecor C., Tyler D.,

RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;

RT "Sequencing of Drosophila melanogaster genome.";

RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.

RP SEQUENCE FROM N.A.

RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,

RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,

RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,

RA Clump M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,

RA Krommiller B., Marshall B., Milburn G., Richter J., Russo S.,

RA Seale S.M.J., Smith E., Shu S., Smutnick F., Whitfield E.,

RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;

RT "Annotation of Drosophila melanogaster genome.";

RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.

RP SEQUENCE FROM N.A.

RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter J.C.;

RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.

RP SEQUENCE FROM N.A.

RA Flybase; Fgn0028707; M2.

DR GO:0003677; P:DNA binding; IEA.

DR GO:0006306; P:DNA methylation; IEA.

DR InterPro: IPR001525; C5 DNA meth.

DR Pfam: PF00145; DNA methylase; 1.

DR PROSITE: PS00095; C5\_MTA5E 2; 1.

SQ SEQUENCE 331 AA; 37870 MW; 6F4BCA669C905FPA CRC64;

QY 6 FTGYSHY 13  
 |||||  
 220 FTGYSHY 227

Db

Query Match 55.6%; Score 40; DB 5; Length 331;  
 Best Local Similarity 75.0%; Pred. No. 73;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

RESULT 12

Q9VKB3 PRELIMINARY; PRT; 345 AA.

AC Q9VKB3; 01-MAY-2000 (TrEMBLrel. 13, Created)

DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE M22 protein (EC 2.1.1.73) (Modification methylase) (Cytosine-specific methyltransferase).

GN M22 OR CG10692.

OS Drosophila melanogaster (fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI\_Taxid=7227;

[1]

RP SEQUENCE FROM N.A.

RX STRAT=Berkeley; PubMed=10731132;

RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,

RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,

RA Brandon R.C., Rogers Y.H., Blazet R.G., Champe M., Pfeiffer B.D.,

RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,

RA Abail J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,

RA Ballew R.M., Baer A., Bayendale J., Bayraktaroglu L., Beasley E.M.,

RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,

RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,

RA Buttle K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,

```

RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Meyers A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Domes K., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glaser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Mei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li X., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Mlshina N.V., Moadary C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclet J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spreading A.C., Stapleton M., Strong M., Sun E.,
RA Svitskay R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang X., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RT Science 287:2185-2195 (2000).
CC -!- CATALYTIC ACTIVITY: S-ADENOSYL-L-METHIONINE + DNA CYTOSINE = S-
CC ADENOSYL-L-HOMOCYSTEINE + DNA 5-METHYLCYTOSINE.
CC -!- SIMILARITY: BELONGS TO THE CS-METHYLTRANSFERASE FAMILY.
DR EMBL: AE003635; AAF5163.1; -.
DR HSSP: 014717; 1G55.
DR FLYBASE: FBgn028707; Mc2.
DR GO: GO:0003677; F:DNA binding; IEA.
DR GO: GO:0008168; F:methyltransferase activity; IEA.
DR GO: GO:0008326; F:site-specific DNA-methyltransferase (cytosi. . .); IEA.
DR GO: GO:0016740; F:transferase activity; IEA.
DR GO: GO:0006306; P:DNA methylation; IEA.
DR InterPro: IPR001525; CS_DNA_meth.
DR Pfam: PF00145; DNA_methylase; 1.
DR PROSITE: PS00095; CS_MTASE_2; 1.
DR Methyltransferase; Transferase.
SQ SEQUENCE 345 AA; 39334 MW; 1BA6F9E65DB12C04 CRC64;

Query Match 55.6%; Score 40; DB 5; Length 345;
Best Local Similarity 75.0%; Pred. No. 76;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 6 FTSGYSHY 13
DB 234 FTKGTYHY 241

RESULT 13
Q9UG67 PRELIMINARY; PRT; 345 AA.
AC Q9UG67;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE DNA (5'-cytosine) methyltransferase homolog (EC 2.1.1.73) (Modification
DE methylase) (Cytosine-specific methyltransferase).
GN MT2 OR DNMT2 OR CG10692.
OS Drosophila melanogaster (Fruit fly).
OC Neoptera; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Ephydroidea; Drosophilidae; Drosophila.
OK NCBI_TaxID=7227;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=99449787; PubMed=10518555;
RA Hung M.S., Karthikeyan N., Huang B., Koo H.-C., Kiger J., Shen C.K.J.;
RT "Drosophila proteins related to vertebrate DNA (5-cytosine)

```

```

RT methyltransferases.";
RT Proc. Natl. Acad. Sci. U.S.A. 96:11940-11945 (1999).
CC -!- CATALYTIC ACTIVITY: S-ADENOSYL-L-METHIONINE + DNA CYTOSINE = S-
CC ADENOSYL-L-HOMOCYSTEINE + DNA 5-METHYLCYTOSINE.
CC -!- SIMILARITY: BELONGS TO THE CS-METHYLTRANSFERASE FAMILY.
DR EMBL: AF185647; AAF03835.1; -.
DR HSSP: 014717; 1G55.
DR FLYBASE: 4245; M.DmeORFAP.
DR GO: GO:0003677; F:DNA binding; IEA.
DR GO: GO:0008168; F:methyltransferase activity; IEA.
DR GO: GO:0008326; F:site-specific DNA-methyltransferase (cytosi. . .); IEA.
DR GO: GO:0016740; F:transferase activity; IEA.
DR GO: GO:0006306; P:DNA methylation; IEA.
DR InterPro: IPR001525; CS_DNA_meth.
DR Pfam: PF00145; DNA_methylase; 1.
DR PROSITE: PS00095; CS_MTASE_2; 1.
DR Methyltransferase; Transferase.
SQ SEQUENCE 345 AA; 39360 MW; 2A8686BE39402788 CRC64;

Query Match 55.6%; Score 40; DB 5; Length 345;
Best Local Similarity 75.0%; Pred. No. 76;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 6 FTSGYSHY 13
DB 234 FTKGTYHY 241

RESULT 14
Q22136 PRELIMINARY; PRT; 459 AA.
ID Q22136;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE T04A8.3 protein.
GN T04A8.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN (1)
RP SEQUENCE FROM N.A.
RA Palmer S.;
RL Submitted (Aug-1994) to the EMBL/GenBank/DBJ databases.
RX MEDLINE=99069613; PubMed=9851916;
SQ SEQUENCE 459 AA; 53253 MW; BD173368D31732A0 CRC64;

Query Match 55.6%; Score 40; DB 5; Length 459;
Best Local Similarity 63.6%; Pred. No. 1e+02;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 DIRKFTSGYS 11
DB 352 DLSKNFTGYS 362

RESULT 15

```



```

Q9K6S0
ID Q9K6S0 PRELIMINARY; PRT; 732 AA.
AC Q9K6S0;
DT 01-OCT-2000 (TRENBLREL. 15, Created)
DT 01-OCT-2000 (TRENBLREL. 15, Last sequence update)
DT 01-JUN-2003 (TRENBLREL. 24, Last annotation update)
DE Glycosyltransferase.
GN BH3657.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=86665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C-125 / JCM 9153;
RX MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
RA Horikoshi K.;
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
RT halodurans and genomic sequence comparison with Bacillus subtilis.";
RL Nucleic Acids Res. 28:4317-4331(2000).
DR EMBL; AP001519; BAB07376.1; -.
DR PIR; A84107; A84107.
DR HSSP; P39621; IQGQ.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0009058; P:biosynthesis; IEA.
DR InterPro; IPR01296; Glyco_trans_1.
DR InterPro; IPR01173; Glyco_trans_2.
DR Pfam; PF00534; Glycos_transf_1; 1.
DR Pfam; PF00535; Glycos_transf_2; 1.
DR KMW transferase; Complete proteome.
SQ SEQUENCE 732 AA; 8415 MW; C483A9BF6B4EAE7D CRC64;

Query March 55.6%; Score 40; DB 16; Length 732;
Best Local Similarity 50.0%; Pred. No. 1.7e+02;
Matches 6; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
QY 2 IRKRTSGSYHY 13
Db 326 VRKEYFDSYHY 337

```

Search completed: October 5, 2004, 08:13:42  
 Job time : 10.8812 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: October 5, 2004, 07:04:10 ; Search time 9.69043 Seconds  
(without alignments)  
371.381 Million cell updates/sec

Title: US-09-805-290A-17  
Perfect score: 72  
Sequence: 1 DIRKRTSGYSHY 13

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_29Jan04:\*  
1: geneseqp1980s:\*  
2: geneseqp1990s:\*  
3: geneseqp2000s:\*  
4: geneseqp2001s:\*  
5: geneseqp2002s:\*  
6: geneseqp2003as:\*  
7: geneseqp2003bs:\*  
8: geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	72	100.0	13	4	AAE10550
2	42	58.3	401	6	ABU26547
3	41	56.9	13	4	AAE10545
4	41	56.9	130	4	AAE10556
5	40	55.6	60	3	AAE01953
6	40	55.6	60	3	AAE27255
7	40	55.6	218	2	AAE24387
8	40	55.6	345	2	ABE65205
9	39	54.2	40	4	AAE17463
10	39	54.2	40	4	ABE36483
11	39	54.2	40	4	ABE31275
12	39	54.2	40	4	ABE21825
13	39	54.2	40	4	AAE69647
14	39	54.2	40	4	AAE57247
15	39	54.2	40	4	ABE51329
16	39	54.2	40	5	ABE39268
17	39	54.2	88	2	AAE27743
18	39	54.2	396	5	ABE39701
19	39	54.2	473	7	ADG96314
20	39	54.2	719	6	ADA34987
21	38	52.8	28	2	AAE22496
22	38	52.8	320	3	AAE21011
23	38	52.8	320	4	AAE39412
24	38	52.8	331	4	AAE41198
25	38	52.8	370	4	ABE67834

26	38	52.8	388	6	ABU20992
27	38	52.8	404	5	ABE52471
28	38	52.8	803	4	AAU12040
29	38	52.8	803	6	ABU25508
30	38	52.8	1238	6	ABU26161
31	37.5	52.1	1327	4	ABE58539
32	37	51.4	13	4	AAE10544
33	37	51.4	71	4	ABG26090
34	37	51.4	130	4	AAE10557
35	37	51.4	225	3	AAE38674
36	37	51.4	237	3	AAE38673
37	37	51.4	252	4	ABE70724
38	37	51.4	265	4	ABE58820
39	37	51.4	292	6	AAE42182
40	37	51.4	364	4	AAE93137
41	37	51.4	375	2	AAW41499
42	37	51.4	375	2	AAW30558
43	37	51.4	375	2	AAW33608
44	37	51.4	630	2	AAE77895
45	37	51.4	630	2	AAW08969

ALIGNMENTS

RESULT 1  
AAE10550  
ID AAE10550 standard; peptide; 13 AA.

AAE10550;  
10-DEC-2001 (first entry)  
Llama species antibody VH CDR3 #19.

Llama antibody; camelid; anorectic; heavy chain variable domain; VH;  
human dietary enzyme inhibitor; medicament; human pancreatic lipase; HPL;  
food; human gastric lipase; HGL; cosmetic control; body weight;  
complementarity determining region 3; CDR3.

Lama sp.  
EP1134231-A1.

19-SEP-2001.

20-FEB-2001; 2001EP-00200703.

14-MAR-2000; 2000EP-00200930.

(UNITL) UNILEVER NV.  
(UNITL) UNILEVER PLC.

Bezemer S, Van De Burg M, De Haard JW, Tareilus E;

WPI; 2001-572718/65.

New antibody or its fragments for inhibiting human dietary enzymes,  
PT useful for cosmetic control of body weight of human beings, comprises  
PT heavy chain variable domain derived from immunoglobulin naturally devoid  
PT of light chains.

Claim 4; Page 29; 37pp; English.

The patent discloses antibodies or their fragments comprising a heavy  
chain variable domain (VH) derived from an immunoglobulin naturally  
devoid of light chains specific for inhibiting human dietary enzymes. The  
antibodies of the invention are useful for the preparation of medicaments  
or food for inhibiting the activity of one or more human dietary enzymes  
especially human pancreatic lipase (HPL) or human gastric lipase (HGL)  
which are useful for the cosmetic control of body weight of human beings.  
The present peptide sequence is a complementarity determining region 3  
(CDR3) of llama species (camelid) antibody VH region



Query Match 56.9%; Score 41; DB 4; Length 13;  
 Best Local Similarity 58.3%; Pred. No. 3.7;  
 Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 2 IRKRFTSGSYSH 13  
 ||:|||||:  
 Db 2 IRRKFTSEYNEY 13

## RESULT 4

AAE10556  
 ID AAE10556 standard; peptide; 130 AA.

XX AC AAE10556;

DT 10-DEC-2001 (first entry)

XX HPL inhibiting VHH fragment, HPL #18 from llama species.

XX llama antibody; camelid; anorectic; heavy chain variable domain; VHH;

XX human dietary enzyme inhibitor; medicament; human pancreatic lipase; HPL;

XX food; human gastric lipase; HGL; cosmetic control; body weight.

OS Lama sp.

XX Key Location/Qualifiers

XX Region 31.35

FT /label= CDR1

FT /note= "Complementarity determining region 1"

FT 50.64

FT /label= CDR2

FT /note= "Complementarity determining region 2"

FT 98.110

FT /label= CDR3

FT /note= "Complementarity determining region 3"

XX EP1134231-A1.

XX 19-SEP-2001.

XX 20-FEB-2001; 2001EP-00200703.

XX 14-MAR-2000; 2000EP-00200930.

XX (UNITL ) UNILEVER NV.

XX (UNITL ) UNILEVER PLC.

XX Bezemer S, Van De Burg M, De Haard JWM, Tarellus E;

XX WPI; 2001-572718/65.

XX New antibody or its fragments for inhibiting human dietary enzymes,

XX useful for cosmetic control of body weight of human beings, comprises

XX heavy chain variable domain derived from immunoglobulin naturally devoid

XX of light chains.

XX Example 2; Page 10; 37pp; English.

XX The patent discloses antibodies or their fragments comprising a heavy

XX chain variable domain (VHH) derived from an immunoglobulin naturally

XX devoid of light chains specific for inhibiting human dietary enzymes. The

XX antibodies of the invention are useful for the preparation of medicaments

XX or food for inhibiting the activity of one or more human dietary enzymes

XX especially human pancreatic lipase (HPL) or human gastric lipase (HGL)

XX which are useful for the cosmetic control of body weight of human beings.

XX The present peptide sequence is HPL inhibiting VHH fragment, HPL #18 from

XX llama (camelid) species

XX Sequence 130 AA;

XX Query Match 56.9%; Score 41; DB 4; Length 130;

XX Best Local Similarity 58.3%; Pred. No. 38;

XX Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 2 IRKRFTSGSYSH 13  
 ||:|||||:  
 Db 99 IRRKFTSEYNEY 110

## RESULT 5

AAE10556  
 ID AAE10556 standard; protein; 60 AA.

XX AC AAE10553;

DT 01-NOV-2001 (first entry)

XX Cyclin dependent kinase inhibitor (CDKI) clone s12.pk0117.h4.

XX Cyclin dependent kinase inhibitor; CDKI; herbicide; cell cycle; soybean;

XX plant growth inhibitor.

XX Glycine max.

XX WO200060087-A2.

XX 12-OCT-2000.

XX 06-APR-2000; 2000WO-US009106.

XX 07-APR-1999; 99US-0128192P.

XX (DUPO ) DU PONT DE NEMOURS & CO E I.

XX Klein TM, Weng Z, Cahoon RE;

XX WPI; 2000-679375/66.

XX N-PSDB; AAN02403.

XX Cyclin dependent kinase inhibitor sequences, useful for identifying

XX herbicides and plant growth inhibitors.

XX Claim 10; Fig 1; 58pp; English.

XX The invention describes a novel isolated polynucleotide comprising a

XX nucleotide sequence encoding one of 17 specific cyclin dependent kinase

XX inhibitor (CDKI) polypeptides, cell cycle regulators involved in control

XX of cell division, growth and death. The nucleotide sequences can be used

XX in a vector to transform a host cell to produce the CDKI polypeptide.

XX They can also be used in methods for selecting and obtaining a nucleic

XX acid sequence that encodes CDKI or affects the level of CDKI expression.

XX The encoded protein can be used in a method for evaluating a compound for

XX its ability to inhibit the activity of a CDKI. The inhibitors can be used

XX as herbicides. They can also be used to inhibit plant growth. The

XX polynucleotide sequences can be used in gene mapping and as genetic

XX markers. The sequence is the soybean CDKI clone s12.pk0117.h4 as

XX described in the method of the invention

XX Sequence 60 AA;

XX Query Match 55.6%; Score 40; DB 3; Length 60;

XX Best Local Similarity 58.3%; Pred. No. 26;

XX Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

XX QY 1 DIRKRFTSGSYSH 12  
 ||:|||||:  
 Db 30 DIKRFTDKYNY 41

XX RESULT 6

XX AAB27255

XX ID AAB27255 standard; protein; 60 AA.

XX AC AAB27255;

XX DT 17-JUN-2001 (first entry)

XX	soybean cyclin-dependent kinase inhibitor #4.
DE	
XX	Soybean, cyclin-dependent kinase inhibitor; cell cycle, cell division;
XX	CDKI; cell growth; herbicide.
KM	
XX	Glycine max.
OS	
XX	W0200060087-A2.
PN	
XX	12-OCT-2000.
PD	
XX	
XX	06-APR-2000; 2000MO-US009106.
PF	
XX	07-APR-1999; 99US-0128192P.
FR	
XX	(DUPO ) DU PONT DE NEMOURS & CO E I.
PA	
XX	Klein TM, Weng Z, Cahoon RE;
P1	
XX	WPI; 2000-679375/66.
XX	
DR	N-PSDB; AAA95289.
XX	
XX	Cyclin dependent kinase inhibitor sequences, useful for identifying
PT	herbicides and plant growth inhibitors.
XX	
PS	Claim 10; Fig 1; 58pp; English.
XX	
CC	The present sequence is the soybean cyclin-dependent kinase inhibitor
CC	(CDKI). Its coding sequence was isolated by searching a soybean seedling
CC	cDNA library for sequences similar to those encoding the CDKI from
CC	Cenopodium rubrum, Caenorhabditis elegans and Arabidopsis thaliana. CDKI
CC	is involved in the cell cycle, and may promote or inhibit cell division
CC	and growth. The protein and its coding sequence are useful in the
CC	production of transgenic plants which produce increased or decreased
CC	amounts of the CDKI protein, in the identification of herbicides, in
CC	genetic and physical mapping and in the isolation of the CDKI gene in
CC	other organisms
XX	
SQ	Sequence 60 AA;
XX	
Query Match	55.6%; Score 40; DB 3; Length 60;
Best Local Similarity	58.3%; Pred. No. 26;
Matches	7; Conservative 3; Mismatches 2; Indels 0; Gaps 0.
QY	1 DIRKFTSGYSH 12
	:     :
	:     :
Db	30 DIQKFTDKXNY 41
XX	
RESULT 7	
AAW24387	
ID	AAW24387 standard; protein; 218 AA.
XX	
AC	AAW24387;
XX	
DT	26-SEP-1997 (first entry)
XX	
DE	Lymantria dispar nuclear polyhedrosis virus 25.7 kDa protein.
XX	
KM	Gypsy moth; nuclear polyhedrosis virus; LdMNPV; insecticide; pesticide;
XX	insect pest; baculovirus; broad host range; virulence; increase.
XX	
OS	Lymantria dispar.
XX	
FN	US5639454-A.
XX	
PD	17-JUN-1997.
XX	
PF	06-APR-1995; 95US-00417822.
XX	
FR	06-APR-1995; 95US-00417822.
XX	

PA	(UNMS ) UNIV MICHIGAN STATE.
XX	
PI	Thiem SM;
XX	
DR	WPI; 1997-332002/30.
DR	N-PSDB; AAT77280.
XX	
PT	Lymantria dispar nuclear polyhedrosis virus p25.7 gene - used to produce
PT	insecticidal nuclear polyhedrosis baculovirus with increased host range.
XX	
PS	Claim 1; Fig 3; 18pp; English.
XX	
CC	This sequence represents a protein of 25.7 kDa encoded by the p25.7 gene
CC	from Lymantria dispar nuclear polyhedrosis virus (LdMNPV). The LdMNPV
CC	p25.7 gene enables recombinant baculoviruses to replicate in cells,
CC	tissue or insect species in which the natural baculovirus isolate cannot
CC	replicate. In addition to extending host range of certain baculoviruses,
CC	the p25.7 gene can also increase virulence against the usual host of the
CC	unaltered virus. For example, insertion of this sequence into Autographa
CC	californica nuclear polyhedrosis virus (AcMNPV) resulted in a significant
CC	reduction in LD50 required to control insects that are susceptible to
CC	infection by unaltered AcMNPV
XX	
SQ	Sequence 218 AA;
Query Match	55.6%; Score 40; DB 2; Length 218;
Best Local Similarity	50.0%; Pred. No. 94;
Matches	6; Conservative 2; Mismatches 4; Indels 0; Gaps 0
Oy	2 IRRKFTSGYSHY 13 : : :
Db	37 VRYRYERGGYHY 48
RESULT 8	
ABB65205	
ID	ABB65205 standard; protein; 345 AA.
XX	
AC	ABB65205;
XX	
DT	26-MAR-2002 (first entry)
XX	
DE	Drosophila melanogaster polypeptide SEQ ID NO 22407.
XX	
KW	Drosophila; developmental biology; cell signalling; insecticide;
KW	pharmaceutical.
XX	
OS	Drosophila melanogaster.
XX	
PN	MO200171042-A2.
XX	
PD	27-SEP-2001.
XX	
PF	23-MAR-2001; 2001WC-US009231.
XX	
PR	23-MAR-2000; 2000US-0191637P.
PR	11-JUL-2000; 2000US-00614150.
XX	
PA	(PEKE ) PE CORP NY.
XX	
PI	Venter JC, Adams M, Li PMD, Myers EW;
XX	
DR	WPI; 2001-656860/75.
DR	N-PSDB; ABD09308.
XX	
PT	New isolated nucleic acid detection reagent for detecting 1000 or more
PT	genes from Drosophila and for elucidating cell signaling and cell-cell
PT	interactions.
XX	
PS	Disclosure; SEQ ID NO 22407; 21pp + Sequence Listing; English.
XX	
CC	The invention relates to an isolated nucleic acid detection reagent
CC	capable of detecting 1000 or more genes from Drosophila. The invention is

CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA  
CC sequences (AB101840-AB16175) and the encoded proteins (AB57737-  
CC AB572072). The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

CC Sequence 345 AA;

Query Match 55.6%; Score 40; DB 4; Length 345;  
Best Local Similarity 75.0%; Pred. No. 1.5e+02;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 6 FTSGYSHY 13  
|||  
234 FTSGYTHY 241

RESULT 9

AA17463  
ID AA17463 standard; protein; 40 AA.

AC AA17463;

DT 12-OCT-2001 (first entry)

DE Peptide #3897 encoded by probe for measuring cervical gene expression.

KM Probe; human; microarray; gene expression; cervical epithelial cell;  
KM cervical cancer.

OS Homo sapiens.

PN WO200157278-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US000670.

PR 04-FEB-2000; 2000US-0180312P.

PR 26-MAY-2000; 2000US-0207456P.

PR 30-JUN-2000; 2000US-00608408.

PR 03-AUG-2000; 2000US-00632366.

PR 21-SEP-2000; 2000US-0234687P.

PR 27-SEP-2000; 2000US-0236359P.

PR 04-OCT-2000; 2000GB-00024263.

PA (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

DR WPI; 2001-488901/53.

FT Human genome-derived single exon nucleic acid probes useful for analyzing

PT gene expression in human cervical epithelial cells.

PS Claim 27; SEQ ID NO 22289; 487bp; English.

XX The present invention relates to human single exon nucleic acid probes

CC (SENPs: see A110068-A128459). The present sequence is a peptide encoded

CC by one such probe. The SENPs are derived from human HeLa cells. The SENPs

CC can be used to produce a single exon microarray, which can be used for

CC measuring human gene expression in a sample derived from human cervical

CC epithelial cells. By measuring gene expression, the probes are therefore

CC useful in grading and/or staging of diseases of the cervix, notably

CC cervical cancer. Note: The sequence data for this patent did not form

CC part of the printed specification, but was obtained in electronic format

CC directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

CC Sequence 40 AA;

Query Match 54.2%; Score 39; DB 4; Length 40;  
Best Local Similarity 87.5%; Pred. No. 25;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 IRRKFTSG 9  
:|||||  
1 VRRKFTSG 8

RESULT 10

ABB36483  
ID ABB36483 standard; peptide; 40 AA.

AC ABB36483;

DT 04-FEB-2002 (first entry)

DE Peptide #3989 encoded by human foetal liver single exon probe.

KM Human; foetal liver; gene expression; single exon nucleic acid probe.

OS Homo sapiens.

PN WO200157277-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US000669.

PR 04-FEB-2000; 2000US-0180312P.

PR 26-MAY-2000; 2000US-0207456P.

PR 30-JUN-2000; 2000US-00608408.

PR 03-AUG-2000; 2000US-00632366.

PR 21-SEP-2000; 2000US-0234687P.

PR 27-SEP-2000; 2000US-0236359P.

PR 04-OCT-2000; 2000GB-00024263.

PA (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

DR WPI; 2001-483447/52.

PT Human genome-derived single exon nucleic acid probes useful for analyzing

PT gene expression in human foetal liver.

PS Claim 27; SEQ ID NO 29118; 639pp + Sequence Listing; English.

XX The invention relates to a single exon nucleic acid probe for measuring

CC human gene expression in a sample derived from human foetal liver. The

CC single exon nucleic acid probes may be used for predicting, measuring and

CC displaying gene expression in samples derived from human foetal liver. The

CC present sequence is a peptide encoded by a single exon nucleic acid probe

CC of the invention. Note: The sequence data for this patent did not form

CC part of the printed specification, but was obtained in electronic format

CC directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

CC Sequence 40 AA;

Query Match 54.2%; Score 39; DB 4; Length 40;  
Best Local Similarity 87.5%; Pred. No. 25;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 IRRKFTSG 9  
:|||||  
1 VRRKFTSG 8

RESULT 11

ABB31275  
ID ABB31275 standard; peptide; 40 AA.

AC ABB31275;

```

XX 01-FEB-2002 (first entry)
XX Peptide #3926 encoded by breast cell single exon nucleic acid probe.
XX Human; microarray; single exon probe; gene expression; breast; disease;
XX cancer.
XX Homo sapiens.
XX WO200157271-A2.
XX 09-AUG-2001.
XX 30-JAN-2001; 2001WO-US000662.
XX 04-FEB-2000; 2000US-0180312P.
XX 26-MAY-2000; 2000US-0207456P.
XX 30-JUN-2000; 2000US-00608408.
XX 03-AUG-2000; 2000US-00632366.
XX 21-SEP-2000; 2000US-0234687P.
XX 27-SEP-2000; 2000US-0236359P.
XX 04-OCT-2000; 2000GB-00024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-496933/54.
XX New spatially-addressable set of single exon nucleic acid probes, useful
XX for measuring gene expression in sample derived from human breast,
XX comprises number of single exon nucleic acid probes.
XX Claim 27; SEQ ID NO 14243; 327bp + Sequence listing; English.
XX The invention relates to a spatially-addressable set of single exon
XX nucleic acid probes for measuring gene expression in a sample derived
XX from human breast and Br 474 cells. The method involves contacting the
XX probes with a collection of detectably labelled nucleic acids derived
XX from mRNA of human breast, and then measuring the label bound to each
XX probe of the microarray. The probes are useful for verifying the
XX expression of regions of genomic DNA predicted to encode proteins. They
XX are useful for gene discovery, and for determining predisposition and/or
XX prognosing breast disease. Gene expression analysis is useful for
XX assessing the toxicity of chemical agents on cells. The microarray of
XX this invention presents a far greater diversity of probes for measuring
XX gene expression, with far less bias than expressed sequence tag
XX microarrays. The method is suitable for rapid production of functional
XX information from genomic sequence. The present sequence is a peptide
XX encoded by a single exon nucleic acid probe of the invention. Note: The
XX sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences
XX Sequence 40 AA;
XX
XX Query Match 54.2%; Score 39; DB 4; Length 40;
XX Best Local Similarity 87.5%; Pred. No. 25;
XX Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
XX
XX 2 IRKRTSG 9
XX :|||||
XX 1 VKRRTSG 8
XX
XX RESULT 12
XX ABB21825
XX ID ABB21825 standard; protein; 40 AA.
XX AC ABB21825;
XX XX
XX PT 23-JAN-2002 (first entry)

```

```

XX Protein #3824 encoded by probe for measuring heart cell gene expression.
XX Human; gene expression; heart; microarray; vascular system;
XX cardiovascular disease; hypertension; cardiac arrhythmia;
XX congenital heart disease.
XX Homo sapiens.
XX WO200157274-A2.
XX 09-AUG-2001.
XX 30-JAN-2001; 2001WO-US000666.
XX 04-FEB-2000; 2000US-0180312P.
XX 26-MAY-2000; 2000US-0207456P.
XX 30-JUN-2000; 2000US-00608408.
XX 03-AUG-2000; 2000US-00632366.
XX 21-SEP-2000; 2000US-0234687P.
XX 27-SEP-2000; 2000US-0236359P.
XX 04-OCT-2000; 2000GB-00024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-48889/53.
XX Single exon nucleic acid probes for analyzing gene expression in human
XX hearts.
XX Claim 15; SEQ ID NO 23595; 530bp; English.
XX The present invention relates to single exon nucleic acid probes for
XX measuring human gene expression in a sample derived from human heart (see
XX ABA21535-ABA41305). The present sequence is a protein encoded by one such
XX probe. The probes may be used for predicting, measuring and displaying
XX gene expression in samples derived from the human heart via microarrays.
XX By measuring gene expression, the probes are useful for predicting,
XX diagnosing, grading, staging, monitoring and prognosing diseases of the
XX human heart and vascular system e.g. cardiovascular disease,
XX hypertension, cardiac arrhythmias and congenital heart disease. Note: The
XX sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences
XX Sequence 40 AA;
XX
XX Query Match 54.2%; Score 39; DB 4; Length 40;
XX Best Local Similarity 87.5%; Pred. No. 25;
XX Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
XX
XX 2 IRKRTSG 9
XX :|||||
XX 1 VKRRTSG 8
XX
XX RESULT 13
XX AAM69647
XX ID AAM69647 standard; protein; 40 AA.
XX AC AAM69647;
XX XX
XX DT 06-NOV-2001 (first entry)
XX DE Human bone marrow expressed probe encoded protein SEQ ID NO: 29953.
XX XX
XX KM Human; bone marrow expressed exon; gene expression analysis; probe;
XX microarray; cancer; leukaemia; lymphoma; myeloma.
XX OS Homo sapiens.
XX

```



```
PN WO200157276-A2.
XX
XX 09-AUG-2001.
PD
XX 30-JAN-2001; 2001WO-US000668.
PF
XX 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00613366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488900/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for analyzing
PT gene expression in human bone marrow.
XX
XX Example 4; SEQ ID NO 29953; 658pp + Sequence Listing; English.
PS
XX CC The present invention provides a number of single exon nucleic acid
XX CC probes which are derived from genomic sequences expressed in the human
XX CC bone marrow. They can be used to measure gene expression in bone marrow
XX CC samples, which may enable the improved diagnosis and treatment of cancers
XX CC such as lymphoma, leukemia and myeloma. The present sequence is a
XX CC protein encoded by one of the probes of the invention
XX
XX SQ Sequence 40 AA;
XX
XX Query Match 54.2%; Score 39; DB 4; Length 40;
XX Best Local Similarity 87.5%; Pred. No. 25;
XX Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 2 IRKRFITSG 9
XX :|||||
XX 1 VRKRFITSG 8
Db
XX
XX RESULT 14
XX AAM57247
XX ID AAM57247 standard; protein; 40 AA.
XX
XX AC AAM57247;
XX
XX DT 05-NOV-2001 (first entry)
XX
XX DE Human brain expressed single exon probe encoded protein SEQ ID NO: 29352.
XX
XX KW Human; brain expressed exon; gene expression analysis; probe; microarray;
XX KW Alzheimer's disease; multiple sclerosis; schizophrenia; epilepsy; cancer.
XX
XX OS Homo sapiens.
XX
XX PN WO200157275-A2.
XX
XX PD 09-AUG-2001.
XX
XX PF 30-JAN-2001; 2001WO-US000667.
XX
XX PR 04-FEB-2000; 2000US-0180312P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-00608408.
XX PR 03-AUG-2000; 2000US-00613366.
XX PR 21-SEP-2000; 2000US-0234687P.
XX PR 27-SEP-2000; 2000US-0236359P.
XX PR 04-OCT-2000; 2000GB-00024263.
XX
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
```

```
XX Penn SG, Hanzel DK, Chen W, Rank DR;
PI WPI; 2001-483446/52.
XX
XX Single exon nucleic acid probes for analyzing gene expression in human
XX PT brains.
XX
XX Example 4; SEQ ID NO 29352; 650pp + Sequence Listing; English.
PS
XX CC The present invention provides a number of single exon nucleic acid
XX CC probes which are derived from genomic sequences expressed in the human
XX CC brain. They can be used to measure gene expression in brain cell samples,
XX CC which may enable the diagnosis and improved treatment of nervous system
XX CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
XX CC epilepsy and cancers. The present sequence is a protein encoded by one of
XX CC the probes of the invention
XX
XX SQ Sequence 40 AA;
XX
XX Query Match 54.2%; Score 39; DB 4; Length 40;
XX Best Local Similarity 87.5%; Pred. No. 25;
XX Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 2 IRKRFITSG 9
XX :|||||
XX 1 VRKRFITSG 8
Db
XX
XX RESULT 15
XX ABG51329
XX ID ABG51329 standard; peptide; 40 AA.
XX
XX AC ABG51329;
XX
XX DT 25-FEB-2003 (first entry)
XX
XX DE Human liver peptide, SEQ ID NO 29977.
XX
XX KW Human; liver; cirrhosis; hyperlipoproteinaemia; hyperlipidaemia;
XX KW hypercholesterolaemia; coronary heart disease.
XX
XX OS Homo sapiens.
XX
XX PN WO200157273-A2.
XX
XX PD 09-AUG-2001.
XX
XX PF 30-JAN-2001; 2001WO-US000664.
XX
XX PR 04-FEB-2000; 2000US-0180312P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-00608408.
XX PR 03-AUG-2000; 2000US-00613366.
XX PR 21-SEP-2000; 2000US-0234687P.
XX PR 27-SEP-2000; 2000US-0236359P.
XX PR 04-OCT-2000; 2000GB-00024263.
XX
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488898/53.
XX
XX PT Human genome-derived single exon nucleic acid probes useful for analyzing
XX PT gene expression in human adult liver.
XX
XX Claim 27; SEQ ID NO 29977; 658pp; English.
XX
XX CC The invention relates to a single exon nucleic acid probe (SENP) (I) for
XX CC measuring human gene expression in a sample derived from human adult
XX CC liver, comprising one of 1109 defined nucleotide sequences given in the
XX CC specification (or complements/ fragments). The probe hybridises at high
```

CC stringency to a nucleic acid molecule expressed in the human adult liver.  
CC (1) may be used for predicting, measuring and displaying gene expression  
CC in samples derived from human adult liver. The genes identified may be  
CC involved in genetic liver diseases such as cirrhosis,  
CC hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which is  
CC associated with coronary heart disease. ABG47348-ABG5930 represent human  
CC liver single exon encoded peptides of the invention. Note: The sequence  
CC information for this patent does not appear in the printed specification  
CC but was obtained in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 40 AA:  
Query Match 54.2%; Score 39; DB 4; Length 40;  
Best Local Similarity 87.5%; Pred. No. 25;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 2 IRKRFTSG 9  
:|||||  
Db 1 VRKRFTSG 8

Search completed: October 5, 2004, 08:00:07  
Job time : 13.8904 secs